

# ABSTRACT

Title of dissertation: APPLICATION OF FLUCTUATION ANALYSIS  
TO CHARACTERIZE MULTI-SCALE NATURE  
AND PREDICTABILITY OF COMPLEX SYSTEMS

Venkat Anurag Setty, Doctor of Philosophy, 2014

Dissertation directed by: Dr. A. Surjalal Sharma  
Department of Astronomy

Complexity is a result of interactions among individual components of a distributed system, each with their own dynamical time scale. Statistical techniques such as fluctuation analysis are used to quantify extent of long range correlations within time series data by estimating a scaling exponent commonly known as Hurst exponent. Data from magnetospheric dynamics, physiology and finance are known to show multi-exponent nature (two exponents in particular) in their fluctuation analysis with a crossover between the power laws. This correlation crossover can be seen due to the statistical approach taken in the analysis of a range of systems with differing dynamical time scales, particularly due to the nature of interactions with one another. We refer to this property as multi-scale nature in the time series data from complex systems. The main contributions of the thesis are as follows:

**Study of crossover in fluctuation analysis of data from magnetosphere, physiology and finance.** We propose an innovative regression scheme, whose mathematical model well describes two exponent nature with an intermediate

crossover regime seen in fluctuation analysis - the Hyperbolic regression. Slopes of the asymptotes of the hyperbola are the Hurst exponents, and, center of the resulting hyperbolic fit is an estimate of the correlation crossover time. It is key to note that in this regression, there are no assumptions made about the crossover time, unlike previous approach to crossover fitting. Different data sets corresponding to different physical processes demonstrate multi-scale nature. However, each data presents a unique challenge to be addressed, as a result of characterization of its scaling crossover. Application of hyperbolic regression on the crossover seen in fluctuation analysis of auroral electrojet index data from magnetosphere resulted in estimation of Hurst exponents before and after the crossover. Also, the correlation crossover time scale is now measured by improved modeling of such data using a stochastic model that demonstrates crossover in fluctuation functions - the OU-Langevin model. Characterization of nature of crossover seen in fluctuation analysis of generalized volatility within financial index data has shown differing nature of financial markets. Such a study would help characterize individual markets utilizing features which were not used previously. Heart rate variability data from healthy patients and patients with congestive heart failure demonstrate differing extent of crossover within the crossover seen in fluctuation functions. This resulted in proposal of a parameter i.e., the extent of crossover parameter that can be used to distinguish patients with the congestive heart failure ailment from normal cases.

**Quantifying predictability of complex systems using Hurst exponents.** Predictability of complex systems suffers due to noise in the data. Long range correlations in noise are seen to cause extreme events or build ups leading to

extreme events in such data. The increased probability of extreme event occurrence makes prediction of resulting time series data difficult. Tail exponent is an exponent resulting out of power laws seen in heavy tailed distributions and is used as a measure of the probability of extreme events in such data. The Hurst exponent which measures the extent of long range correlations using fluctuation analysis has a known relationship with the tail exponent through Taqqu's theorem. Fluctuation analysis of time series data is oftentimes complicated due to existence of trends in the time series data. Dynamical features in data commonly reflect as trends, and, as a result nonlinear dynamical time series prediction techniques can be used to measure these trends. We refer to this method as Fluctuation Analysis after Trend Elimination (FATE) and apply it to data from space weather and finance. Hurst exponent estimated from FATE and its relationship with tail exponent measured by a commonly used Hill estimator technique are shown. Insufficient data sizes limit our ability to robustly estimate the tail exponent from observational data as extreme events are usually rare. This forms the motivation to use Hurst exponent obtained from FATE as a measure of predictability of complex systems demonstrated by application on auroral electrojet index data. Conversely, Hurst exponent can be used to quantify the ability of a technique to predict time series data as demonstrated by application of FATE.

Application of Fluctuation Analysis to Characterize Multi-scale  
Nature and Predictability of Complex Systems

by

Venkat Anurag Setty

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2014

Advisory Committee:

Dr. A. Surjalal Sharma, Chair/Advisor

Dr. Rajarshi Roy

Dr. Eugenia Kalnay

Dr. Dennis Papadopoulos

Dr. Brian R. Hunt

© Copyright by  
Venkat Anurag Setty  
2014

## Acknowledgments

I would like to express my gratitude to all the people who have made this thesis possible and my graduate student life memorable.

First and the foremost, I would like to convey heartfelt thanks to my advisor Dr. A. Surjalal Sharma, for providing me with an opportunity to work on challenging and extremely interesting scientific questions over the past three years. His guidance was paramount to my work and finishing of thesis. More importantly, his dedication and determination towards my progress kept me grounded and focused. My association with him would be cherished and the lessons learned valued for a lifetime.

Thanks are due to Professor Rajarshi Roy (Dean's representative), Professor Dennis Papadopoulos , Professor Eugenia Kalnay and Professor Brian Hunt for agreeing to serve on my thesis committee and for sparing their invaluable time reviewing the manuscript. I would like to thank Professor Michael Coplan and Debbie Jenkins at the Chemical Physics program for their constant help and support.

I owe deepest thanks to my family who have always stood by me, guided me through my career, and have pulled me through against impossible odds at times.

My colleagues at the Space Plasma Physics and theoretical Nonlinear Physics groups have enriched my graduate life in many ways and deserve a special mention. Dr. Francesco Sorrentino, Dr. Bhargava Ravoori and Dr. Adam Cohen provided me with much needed support and taught me many scientific tools during my first research work on Adaptive Synchronization. Dr. Nicholas Mecholsky was my office

mate during my early research years and his support and kindness towards a young graduate student is never forgotten. Aram Vartanyan, Chris Najmi and Prabin Adhikari have been my go to support group for general counsel on issues regarding graduate student life and scientific questions over the last few years of my graduate work. I wish them very best in their respective scientific endeavors.

I would like to acknowledge financial support from the Chemical Physics program, Department of Chemistry and grants from NSF.

# Table of Contents

List of Figures	vi
List of Abbreviations	x
1 Introduction	1
1.1 An overview . . . . .	1
1.2 Complex multi-scale systems . . . . .	9
1.3 Long Range Correlations and Hurst exponent . . . . .	18
1.4 Crossover phenomenon in fluctuation functions and probability distributions . . . . .	22
1.5 Nonlinear time series prediction . . . . .	24
1.6 Outline of Thesis . . . . .	33
2 Characterizing Detrended Fluctuation Analysis of Multifractional Brownian Motion	34
2.1 Overview . . . . .	34
2.2 Fractional and Multifractional Brownian Motion . . . . .	38
2.3 Detrended Fluctuation Analysis . . . . .	40
2.4 Estimation of time averaged Hurst exponent by DFA . . . . .	44
2.5 Analysis of DFA performance . . . . .	47
2.6 Time varying Hurst exponents in AL index data . . . . .	54
2.7 Conclusions . . . . .	56
3 Modeling crossover phenomenon in Detrended Fluctuation Analysis of Auroral Electrojet index data	58
3.1 Overview . . . . .	58
3.2 Fluctuation Analysis and scaling crossovers in AL index data . . . . .	61
3.3 Theoretical models for crossover phenomena . . . . .	67
3.4 Stochastic modeling of AL index and estimation of crossover time . . . . .	74
3.5 Conclusions . . . . .	76



4	Varying nature of crossover phenomenon across global financial markets	78
4.1	Overview . . . . .	78
4.2	Scaling crossover in financial indices . . . . .	80
4.2.1	Crossover in financial data and previous results . . . . .	81
4.2.2	DFA of financial indices and crossover analysis . . . . .	82
4.2.3	Convexity and concavity in LRC . . . . .	86
4.3	Conclusions . . . . .	91
5	Scaling crossover in heart rate variability : Extent of crossover analysis	93
5.1	Overview . . . . .	93
5.2	Scaling crossover in HRV . . . . .	95
5.2.1	Crossover in RR interval data and previous results . . . . .	96
5.2.2	Crossover analysis for NSR cases . . . . .	98
5.2.3	Crossover analysis for CHF cases . . . . .	99
5.2.4	Extent of crossover analysis . . . . .	100
5.3	Conclusions . . . . .	104
6	Characterizing predictability using fluctuation analysis	106
6.1	Overview . . . . .	106
6.2	Technique demonstrated on a theoretical model . . . . .	108
6.3	Tail exponent estimation using Hill estimator . . . . .	110
6.4	FATE and characterization of tail exponent . . . . .	110
6.4.1	AL index data . . . . .	111
6.4.2	Microsoft stock price data . . . . .	117
6.5	Conclusions . . . . .	121
7	Summary	123
	Bibliography	129

## List of Figures

1.1	Illustration demonstrating scaling crossover in power-laws, particularly in fluctuation functions that measure the extent of long range correlations within time. Plot (a) represents a mono-scale system where as plots (b) and (c) represent two-scale systems with increasing or decreasing long range correlations, respectively. . . . .	5
1.2	Illustration showing the comparison between normal distribution and fat-tailed distributions (From <a href="http://www.fattails.ca/distribution.html">http://www.fattails.ca/distribution.html</a> ). We see that the probability of occurrence of events of large magnitudes remains finite in fat-tailed distribution, however the probability of occurrence for similar events approaches zero for a normal distribution. . . . .	7
1.3	Schematic showing the outline of ideas behind the thesis work. . . . .	8
1.4	A three dimensional sketch of the magnetosphere and the associated currents, fields and plasma regions [Kivelson and Russell 1995]. . . . .	11
1.5	The sky lit green due to aurora at Hamary in Nordland, Norway (From <a href="http://www.telegraph.co.uk/earth/earthpicturegalleries/9269571/Spectacular-displays-of-the-northern-lights-or-aurora-borealis-in-northern-Norway.html">http://www.telegraph.co.uk/earth/earthpicturegalleries/9269571/Spectacular-displays-of-the-northern-lights-or-aurora-borealis-in-northern-Norway.html</a> ). . . . .	12
1.6	A normal lead II ECG wave progression for a single cycle (From <a href="http://www.sci.utah.edu/macleod/bioen/be6000/labnotes/bp/descrip.html">http://www.sci.utah.edu/macleod/bioen/be6000/labnotes/bp/descrip.html</a> ). . . . .	15
1.7	Probability distribution of AL index data from 1980-2013 (in blue). Comparison between observed probability and normal distribution (in red) shows the evidence of heavy tailed nature of distribution. . . . .	23
1.8	Estimation of time delay and embedding dimension parameters for phase space reconstruction of Lorenz attractor. . . . .	28
1.9	Prediction of $x(t)$ of Lorenz attractor using weighted mean field filter technique. . . . .	32
2.1	Application of Detrended Fluctuation Analysis with quadratic detrending (DFA2) to fractional Gaussian noise with Hurst exponents, $H=0.3, 0.5$ & $0.7$ . . . . .	41

2.2	Simulated mBm data with various forms of $H(t)$ with the incremental process, mGn shown in upper panel and the corresponding mBm in bottom panel. In (a), $H_1(t) = c = 0.75$ . In (b), $H_2(t) = at + b$ with $a = 10^{-5}$ and $b = 0.7$ i.e., $\langle H \rangle_{act} = 0.75$ and $R = 0.1$ . In (c), $H_3(t) = H_2(t) + \epsilon_t$ with $\epsilon_t$ drawn from normal distribution, $\mathcal{N}(0, \sigma^2)$ such that $\sigma = \langle H \rangle_{act} \times 10^{-3}$ . It is interesting to note the differences in variability in mGn resulting in differences in LRC natures of mBm.	45
2.3	Dependence of $\langle H \rangle_{est}$ on the polynomial order ( $p$ ) of DFAP used. The actual average values ( $\langle H \rangle_{act}$ ) are shown in red. $R = 0.06$ for all $H(t)$ used and $T = 2^{13} = 8192$ . We see $p = 1$ has higher $EE$ compared to other cases.	48
2.4	Dependence of $\langle H \rangle_{est}$ on the length of data used for $\langle H \rangle_{act} = 0.75$ and $R = 0.06$ . It is interesting to note that while increasing data size has no predominant effect on $EE$ , it does improve the quality of performance or decrease the $SE$ as shown by error bars.	50
2.5	Contour map showing the dependence of relative accuracy of estimation (legend shown in color map on the right) on $\langle H \rangle_{act}$ and $R$ . We see interesting features emerging from this picture in that data with $\langle H \rangle_{act} \geq 0.75$ have predominantly lower $EE$ compared to other cases, and, $R = 0.06$ serves as a transition point where $EE$ increases rapidly beyond 2% approaching 5% near $R = 0.1$ .	51
2.6	Dependence of $\langle H \rangle_{act}$ on the strength of Gaussian noise ( $\mathcal{N}(0, \sigma^2)$ ) added to $H(t)$ , represented here by $\kappa = \sigma \times \langle H \rangle_{est}$ . We note how rapidly $EE$ increases, well beyond the 5% limit (as shown in green line) as the strength increases beyond $10^{-3}$ .	54
2.7	Time varying Hurst exponent picture of AL index data from 2001-2012 shown as weekly Hurst exponents.	55
3.1	Power spectrum of 5 minute AE index from 1978-1980, we see a break/crossover in the spectrum at the frequency corresponding to a time of about 5 hours [Tsurutani et al. 1990].	59
3.2	DFA of AL index- Fluctuation plots for AL index of 10 weeks in length averaged over 100 ensembles. DFA for various orders of detrending polynomial, $n = 1, 2, 3, 4, 5$ & $6$ is shown to illustrate the dependence of fluctuation plots on $n$ .	63
3.3	Hurst exponent estimation from FA, DFA and cross over point detection using Hyperbolic least square fitting for AL data. FA fails to estimate the correct Hurst exponent before the crossover due to the non-stationary nature of AL data. DFA1 and DFA2 do a better job but show different crossover points.	65
3.4	DFA1 of AL (in blue) and solar wind data (in red). We see that the long term Hurst exponent of AL is equal to the short term Hurst exponent of solar wind data.	67
3.5	DFA1 of sine and its successive differences.	70
3.6	DFA1 of Lorenz and its successive differences.	71

3.7	DFA1 of OU-Langevin and its successive differences. . . . .	72
3.8	DFA1 of successive differences of AL index data. . . . .	73
3.9	Variance of AL estimated from multiple trajectory ensemble (MTE) approach (in blue) and the theoretical curve from an OU-Langevin process with similar crossover time of 265 minutes or 4 hours 25 minutes (in red). . . . .	75
4.1	Daily closing prices for DJIA, NASDAQ, SSE, N225 and FTSE for 5990 days of trading days till April 27 <sup>th</sup> , 2014. . . . .	80
4.2	DFA of generalized volatility of SSE at various power exponents [Shi-Hao 2009] in order of increasing power exponents (ordered top to bottom) shown in (a). Also, the DFA exponent shows the separation of scaling across crossover and is shown in relation with power exponent in (b). . . . .	82
4.3	DFAn and FA of DJIA generalized volatility for its complete data length (1986-2014). Note: window size $\Delta T$ is in days. . . . .	83
4.4	DFAn and FA of DJIA generalized volatility from 1990-2014. Note: window size $\Delta T$ is in days. . . . .	84
4.5	DFAn and FA of NASDAQ generalized volatility. Note: window size $\Delta T$ is in days. . . . .	85
4.6	DFAn and FA of SSE generalized volatility. Note: window size $\Delta T$ is in days. . . . .	86
4.7	DFAn and FA of N225 generalized volatility. Note: window size $\Delta T$ is in days. . . . .	87
4.8	DFAn and FA of FTSE generalized volatility. Note: window size $\Delta T$ is in days. . . . .	88
4.9	DFA2 of generalized volatility of all indices. . . . .	89
4.10	Comparison between concave and convex LRC seen in two theoretical models. Plots correspond to DFA1 of fGn+fBm and OU-Langevin models with a line representing $H = 1$ shown for reference. . . . .	90
4.11	Concavity versus convexity seen in DFA2 of generalized volatility of various indices. The points show the separation between long time and short time DFA exponents ( $\alpha_L$ and $\alpha_S$ respectively) in relation to dotted line that represents absence of crossover. . . . .	91
5.1	Interbeat interval data for 2000 beats of normal healthy subject (upper panel), and, patient with CHF (lower panel). . . . .	96
5.2	Result from previous work [Peng et al. 1998] showing DFA of RR interval data from normal (NSR) and congestive heart failure (CHF) cases. We see that the NSR case has a decreasing LRC while CHF case has increasing LRC with time respectively. The window lengths here are in units of beat numbers. . . . .	97

5.3	Hyperbolic regression and asymptotic fit of fluctuation functions from DFA1 of RR interval data of a NSR subject with DFA1 in blue and regression fit in red. The regression identifies that the two asymptotes are parallel or that the linear fit is the best fit. . . . .	99
5.4	Hyperbolic regression and asymptotic fit of fluctuation functions from DFA1 of RR interval data of a CHF patient. The angle between the asymptotes (red and blue dashed lines) in this particular case is measured to be $\theta_c = 22.31^\circ$ . . . . .	100
5.5	EOC for NSR and CHF patients. The EOC parameter for NSR is shown in circles and for CHF in crosses. We see that $\theta_c$ for most NSR patients is below $5^\circ$ and for most CHF patients is at or above $10^\circ$ . . .	101
5.6	Scatter plot showing $\alpha_s$ and $\alpha_l$ for short and long terms respectively for NSR and CHF patients with dashed line representing $\alpha_s = \alpha_l$ . We see that the $\alpha_s$ is very close to $\alpha_l$ for most NSR patients and the difference in CHF patients is apparent. . . . .	102
6.1	FATE applied to test data with fGn ( $H = 0.86$ ) riding on top of sinusoidal trend. We see that FA of detrended data retrieves most of the scaling properties of the fGn with explicit detrending of sine trend using time series predictions. . . . .	109
6.2	AL (actual in blue) and predicted (in red, black and green) for autonomous, combined and non-autonomous methods of predictions . .	112
6.3	FATE applied to AL data with several prediction methods. We see that non-autonomous is not very different from FA of AL data. Also, autonomous predictions result in $H = 0.5$ with most of crossover effects removed. . . . .	113
6.4	Local slope of FATE on AL. We see that the non-autonomous model (in black) retains most of the crossover effects shown in AL data (in blue). . . . .	114
6.5	MSFT closing stock price (actual in blue) and predicted (in red) for 1000 trading days ending on September 11th, 2014. . . . .	118
6.6	FATE applied to MSFT stock price data. The DFA2 of MSFT yields $\alpha = 1.41$ or $H = 0.41$ , where as FATE yields $H = 0.51$ . . . . .	119

## List of Abbreviations

AE	Auroral Electrojet index
CHF	Congestive Heart Failure
DFA	Detrended Fluctuation Analysis
DJIA	Dow Jones Industrial Average
ECG	Electrocardiogram
EMH	Efficient Market Hypothesis
EOC	Extent Of Crossover
FA	Fluctuation Analysis
FATE	Fluctuation Analysis after Trend Elimination
fBm	fractional Brownian motion
fGn	fractional Gaussian noise
FTSE	Financial Times Stock Exchange
HRV	Heart Rate Variability
IMF	Interplanetary Magnetic Field
LRC	Long Range Correlations
mBm	multifractional Brownian motion
mGn	multifractional Gaussian noise
N225	Nikkei 225
NASDAQ	National Association of Securities Dealers Automated Quotations
NMSE	Normalized Mean Square Error
NSR	Normal Sinus Rhythm
PSD	Power Spectral Density
SSE	Shanghai Stock Exchange

## Chapter 1: Introduction

### 1.1 An overview

Complex systems involves study of how multiple interacting components or parts results in collective behavior of the system and their relationship with respective environments. Such systems are modeled using a combination of approaches from statistical physics, nonlinear dynamics and information theory. Their complexity is due to the nonlinear and seemingly unpredictable outcome of the aggregate activity commonly measured as an output time series. Due to diverse natures of the component processes, a formal structure for modeling such systems does not exist, complicating the simulation of such systems.

Introduction to such complex systems is provided in Section 1.2. It is to be noted that the statistical analysis techniques which would be introduced are applied to data from complex systems in this thesis. Hence, all the ideas are presented within the context of complex systems and in scope of ideas relating to complexity and fractality of such systems.

We now present the central ideas behind the thesis work in this section as an overview to better prepare for the in-depth analysis that is to follow later.

## Power laws: Basic ideas

Power laws occur are seen in certain statistical functions over a range of scales pertaining to input argument of the function ( $x$  in Equation 1.1). Mathematically, a function  $f(x)$  is said to exhibit power law (with respect to  $x$ ) if:

$$f(x) = kx^a \tag{1.1}$$

where,  $k$  and  $a$  are constants. Parameter  $a$  is generally referred to as the scaling exponent. Such power laws are seen in a wide range of statistical measures in various physical systems. The topic of power-law analysis has been extensively reviewed in various review articles [Mitzenmacher 2004, Newman 2005, Sornette 2006]. Power-laws are often seen in probability distributions of various physical phenomena such as solar flares, Earthquake data, internet data, etc. [Clauset et al. 2009].

Long range correlations broadly refer to phenomena where certain statistical or physical measures exhibit correlation over scales that characterize typical behavior. One such example can be seen in the phenomenon of critical opalescence at the liquid-gas critical point. Due to breakdown of distinction between a liquid and a gas at the critical point where due to special combination of pressure and temperature, the liquid is sufficiently hot and compressed that it almost becomes a gas. Such transitions between liquid and gas at the critical point are second-order phase transitions and the density fluctuations in the liquid can be seen at all possible wavelengths leading to the scale free nature.

Such long range correlations seen in the critical regime complicate the anal-



ysis of phase transitions owing to violation of mean-field approaches which were widely used in such analysis [Guggenheim 1945]. Development of renormalization-group [Wilson 1971] helped develop a method for studying systems with scale free nature. There are many physical measures which exhibit long range correlations. However, memory within data results in statistical similarity over several time or length scales [Beran 1994].

Long range correlations in a time series results in the scale free property of the variance estimated at various time scales. This yields a scaling exponent widely known as Hurst exponent, which is most commonly estimated using a class of techniques known as Fluctuation Analysis. We are primarily interested in power-laws seen in fluctuation functions measured from Fluctuation Analysis and the resulting exponent known as Hurst exponent which is used as measure of extent of long range correlations.

We now discuss the phenomenon of scaling crossovers seen in fluctuation analysis and establish a connection between crossovers and multi-scale/multi-fractal nature of complex data in particular.

## Multi-scale nature and Scaling crossovers

Dynamically, a multi-scale system can be defined as one that incorporates processes at different spatial and temporal scales. Heterogeneity among the interacting parts results in multiple scales as seen spatio-temporally in the time series. Modeling of such systems needs an understanding of the role played by different underlying

components at various scales. Hence, study of various scales existing within the time series data is imperative in identifying and applying the relevant model for the data.

Due to the complex nature of interactions, such dynamical time scales are not always distinguishable. Statistical correlations measured from fluctuation analysis include correlations of all types-linear and nonlinear. Scaling crossovers refer to the break in scaling function of interest i.e., fluctuation functions which are a measure of extent of long range correlations. Particularly, existence of scaling crossovers implies violation of scale free nature in fluctuation functions. As an example, data from magnetospheric dynamics demonstrate interactions among the many modes corresponding to various plasma processes. Although the dynamical time scales corresponding to individual processes are hard to identify, the statistical correlations between these processes as seen in data reflect as a power law in fluctuation functions or power spectral density. Violation of this power law through a crossover implies changing correlation time scales. We define and refer to this as multi-scale nature of data throughout the thesis. As the scaling crossover is easily seen in statistical measures like fluctuation functions, it helps us identify correlation time scales. The crossover time scale is a measure of change in nature of correlations within the data.

As we introduce later in Chapter 2, Detrended Fluctuation Analysis (DFA) is one estimation technique widely used for fluctuation analysis of non-stationary data. Non-stationarity in data can result from trends in data. Such trends are either built into the system in a additive fashion i.e.,  $\text{data}=\text{trend}+\text{variability}$  (separable trends in general) or multiplicative fashion where variability affects the evolution,

e.g., noise added to differential equation describing such systems. These effects and theoretical models explaining them are described in detail in Chapter 4.

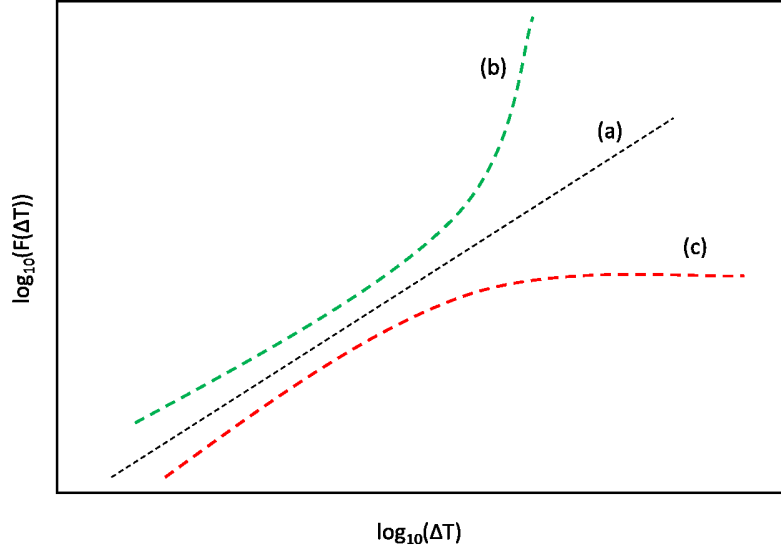


Figure 1.1: Illustration demonstrating scaling crossover in power-laws, particularly in fluctuation functions that measure the extent of long range correlations within time. Plot (a) represents a mono-scale system where as plots (b) and (c) represent two-scale systems with increasing or decreasing long range correlations, respectively.

Data from complex systems oftentimes have processes at different time scales. Most commonly seen are two-scale systems with low frequency and high frequency processes i.e., trend and variability respectively. Processes that result in high frequency of power spectrum in such systems contribute to scaling at short times and those at low frequency contribute to scaling at long times. As an example, we present an illustration in Figure 1.1 where we present typical fluctuation functions from various kinds of systems. Power law clearly exists in system that shows linearity in fluctuation functions such as in plot (a).

Scaling crossovers in two-scaled systems can be of two different natures as seen in plots (b) and (c) in Figure 1.1 i.e., increased scaling exponents through crossover or decreasing exponents through the crossover point, respectively. Data with crossover nature seen in plot (b) can be readily modeled using additive models i.e.,  $\text{data}=\text{trend}+\text{variability}$  because fluctuation functions are additive resulting in fluctuation function of the data being the sum of fluctuation functions of trend and the variability. However, crossovers seen in plot (c) need a complicated model which has noise added to dynamics.

## Motivation for crossover analysis

Irrespective of the nature of crossover, one needs to separate the different scaling behaviors and one way is to identify the point of crossover and measure the exponents before and after the crossover point. This is the central theme of much of the work in this thesis.

The existence of multiple scaling exponents implies a violation of the scale free nature, often seen in fluctuation analysis. The simplest form of multiple scales would consist of two correlation scales with an intermediate crossover regime. This separation of scaling features at various time scales can be used as an evidence of multi-scale nature in observed data of complex systems. Identification and characterization of such a crossover is essential for understanding the nature of underlying physical processes which may result in a better model for such systems. Proposal of a novel scheme to study the crossover phenomena in various real world data em-

bodying complexity and subsequent results from its application constitute one of the two primary subject matter of this thesis.

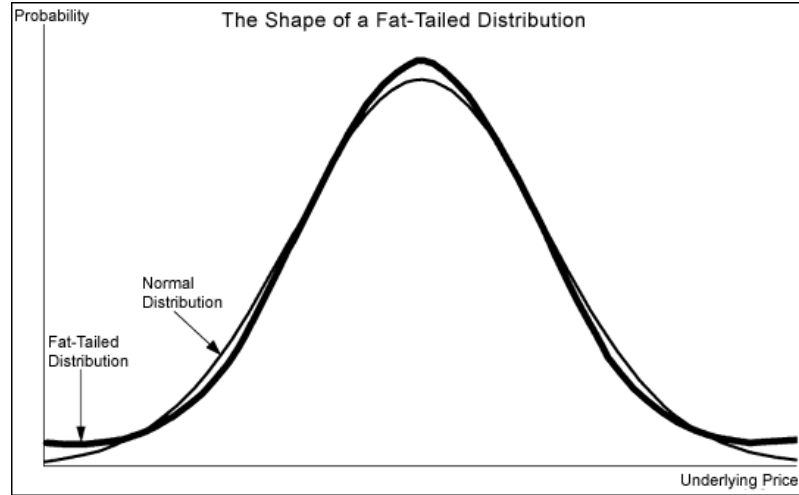


Figure 1.2: Illustration showing the comparison between normal distribution and fat-tailed distributions (From <http://www.fattails.ca/distribution.html>). We see that the probability of occurrence of events of large magnitudes remains finite in fat-tailed distribution, however the probability of occurrence for similar events approaches zero for a normal distribution.

Furthermore, we present an additional case that fat-tailed probability distributions (or leptokurtic distributions) that occur due to existence of so called Black-swan events in data can be interpreted as crossover in the probability distribution. This is a new way of studying fat-tails. Figure 1.2 shows an illustration that shows the difference between normal and fat-tailed distributions. Clearly the fat tails occur due to infinite variance in distribution function and it can be seen as scaling crossover from normal distribution before crossover to a power-law fat tail after the crossover.

# Characterization of predictability

Existence of trends often times complicate estimation of Hurst exponent in time series data. These trends are seen forming the dynamical structure within the data. Processes at short time scales (high frequency component in the power spectral density) of time series decrease the predictability of such time series. These processes are commonly seen as fluctuations or called as noise within the data. Understanding the long range correlations among fluctuations helps us estimate the tail exponent or a measure of extreme event probability in such fluctuations. Hence, the other focus area of this thesis work is a better trend estimator and using Hurst exponents of the detrended time series to characterize the predictability.

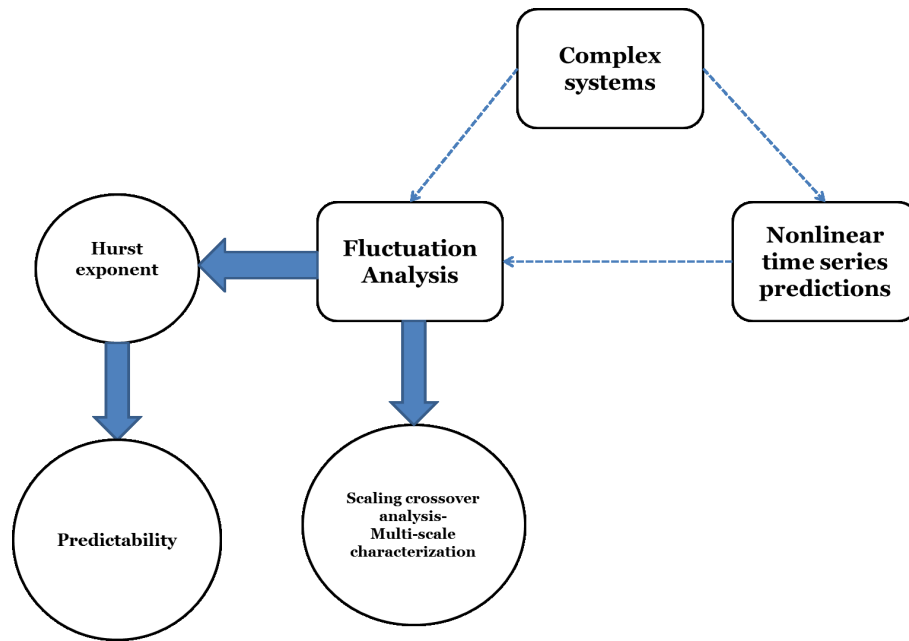


Figure 1.3: Schematic showing the outline of ideas behind the thesis work.

The schematic in Figure 1.3 shows the organization of the ideas in the thesis. Complex data is analyzed using fluctuation analysis in two ways-crossover analysis and study of predictability using Hurst exponent. The results lead to multi-scale characterization and understanding predictability of such complex data.

In the following sections, ideas on complexity, introduction to data and techniques used are presented.

## 1.2 Complex multi-scale systems

The scientific study of complex systems is relatively young in comparison to conventional fields of science with simple system assumptions, such as from physics and chemistry. Among the key contributions to the study of complex systems is the discovery of chaos in deterministic systems, a feature of certain dynamical systems that is strongly related to nonlinearity. Study of neural networks was also integral in advancing the mathematics needed to study complex systems. The science of complex systems mainly deals with the study of how individual interactions give rise to collective and emergent behaviors and the interaction of the aggregate system with its surroundings [Bar-Yam 2002].

Multi-scale nature were discovered in data from various real world systems including those from climate [Goodin et al. 2002], finance [LeBaron 2006], space weather [Tsurutani et al. 1990] etc., Research work leading to the current thesis first started with a study of time series prediction techniques applied on space weather data particularly the auroral electrojet index data corresponding to magnetospheric

dynamics. Scaling crossover was seen in power spectral density of auroral electrojet index data. Fluctuation analysis is an improvement over power spectral density. Our work identified similar scaling crossover in fluctuation analysis as seen from power spectral density. Hence, we first introduce magnetospheric dynamics and auroral electrojet index data in this Section.

Subsequently, existence of two correlation time scales with a distinguishable crossover was found in data from generalized volatility of financial index in finance and heart beat variability from physiology. Identification and characterization of crossovers helped in better modeling of such data. However, problems remained in the method of characterization of scaling crossover in fluctuation analysis. Our work focuses on addressing this challenge and studying the following results. Hence, we next introduce data corresponding to financial indices and heart beat variability in this Section.

It is to be noted that each data set presents a unique challenge or a problem to be addressed with respect to its multi-scale property. While existence of a crossover in scaling is the unifying feature between all the data, our work on characterization of this feature yielded new and interesting results in each case.

## Magnetospheric dynamics

Space weather usually refers to the conditions in the near-Earth space that can potentially impact the satellites and other technological infrastructure here on and around the Earth. Bursts of plasma and magnetic field structures from the Sun's



atmosphere known as coronal mass ejections (CME) together with solar flares can cause space weather effects on the Earth. The plasma matter from the Sun during CMEs travels towards the Earth incorporating various time and spatial scales in its course as the solar wind. Due to mixing of spatio-temporal scales, the data corresponding to solar wind is known to be turbulent.

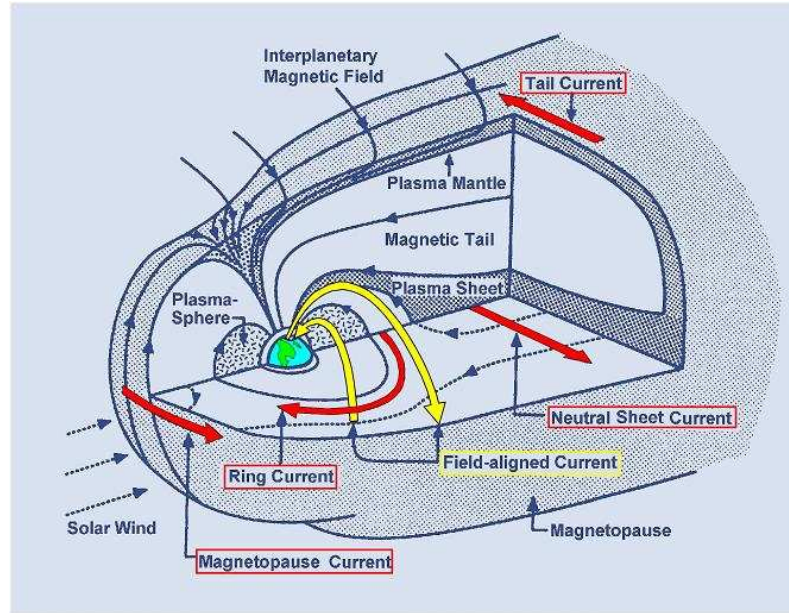


Figure 1.4: A three dimensional sketch of the magnetosphere and the associated currents, fields and plasma regions [Kivelson and Russell 1995].

As the solar wind approaches a planet with well developed magnetic field structure around it, such as the Earth, it interacts with its atmosphere. The region around a planet is the magnetosphere which acts as a shield deflecting the incoming bombardment of solar wind around rather than reaching the planet's surface. The structure of Earth's magnetosphere with respect to incoming solar wind is shown in Figure 1.4. Important components of magnetosphere include bow shock, magne-

tosheath, magnetopause, magnetosphere, northern tail lobe, southern tail lobe and plasmasphere [Ratcliffe 1972].



Figure 1.5: The sky lit green due to aurora at Hamar in Nordland, Norway (From <http://www.telegraph.co.uk/earth/earthpicturegalleries/9269571/Spectacular-displays-of-the-northern-lights-or-aurora-borealis-in-northern-Norway.html>).

The interaction of the solar wind with the magnetosphere and ionosphere at higher latitudes leads to the phenomenon of *aurora borealis*, commonly known as the aurora shown in Figure 1.5. The orientation of interplanetary magnetic field (IMF) of the solar wind expectantly controls the configuration of Earth's magnetosphere. Solar wind-magnetosphere coupling is enhanced when this IMF turns southward as a portion of the transferred energy accumulates in the magnetotail and explosively released in a magnetospheric substorm. Magnetosphere remains highly dynamic under these conditions and the magnetospheric substorms result in elementary disturbances with a typical time scale of an hour. When the IMF remains southward for

an extended period of time, the energy transfer from the solar wind into magnetosphere continues further and the ring current grows under the influence of the solar wind variations leading to geospace storms. These are characterized by a strong decrease in the surface magnetic field. The extremes of space weather occur when intense solar activity is strongly coupled to the solar wind-magnetosphere-ionosphere system during geospace storms.

Auroral Electrojet (AE) refers to the large horizontal currents that flow in D and E layers of the auroral ionosphere. These auroral electrojet currents are known for their strength and persistence due to increased conductivity of the auroral ionosphere. AE was introduced as a measure of the global electrojet activity [Davis and Sugiura 1966]. This index is derived from geomagnetic variations in the horizontal component observed at selected observatories along the auroral zone and is a direct measure of the substorm strength. AE is computed at 1 minute resolution after removal of an average over five quietest days in a month at each observatory. The auroral electrojet indices comprise of AU- the upper bound of data from all the stations, AL-the lower bound, AO-the average between AU and AL, and finally the AE-the difference between AU and AL indices respectively. The AL index is known to be particularly sensitive to the input solar wind driver and is used as the data for our research. This index is obtained from geomagnetic data service of World Data Center-Kyoto: <http://wdc.kugi.kyoto-u.ac.jp/aedir/index.html>.

The strength of solar wind is measured as the strength of electric field of the solar wind flow and thus reflected in the vector cross product between solar wind velocity and IMF strength. These parameters are measured from the

bulk speed ( $v$  in km/s) and strength of IMF in the north-south direction ( $B_z$  in nT) of solar wind as their product- $vB_z$ . Data pertaining to these parameters are obtained real time from measurements by Advanced Composite Explorer (ACE) space craft that is located at the L1 point in the orbit between Earth and the sun: <http://www.swpc.noaa.gov/ftplib/lists/ace/>.

Study of space weather has been not only necessary but also important due to the fact that modern day technological infrastructure is risk prone due to extreme events [Baker et al. 1998]. Because of the non-equilibrium nature of the statistics of such systems, we rely on the framework provided by nonlinear dynamics and complexity science to study them [Baker et al. 1990, Sharma 1995, Vassiliadis et al. 1990]. The dynamics of magnetosphere can be described by a trajectory in an appropriately defined state space which then can be used to describe the system completely. Applying techniques from dynamical systems theory, we can obtain these trajectories from observational data, model its internal dynamics and develop predictions based on the expected evolution of such trajectories, as shown in the Section 1.4.

## Heart rate variability

Heart rate variability (HRV) is the phenomenon of variation in time intervals between successive heartbeats. Electrocardiography (ECG) is the most commonly used method of measuring the HRV data. This method involves recording of electrical activity of the heart using electrodes attached to the surface of the skin. The

measured data is most commonly displayed/recorded externally on a digital interface/recording device.

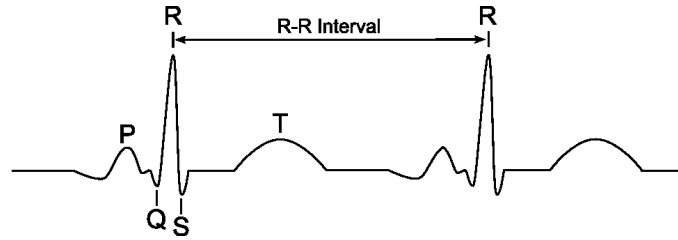


Figure 1.6: A normal lead II ECG wave progression for a single cycle (From <http://www.sci.utah.edu/macleod/bioen/be6000/labnotes/bp/descrip.html>).

In Figure 1.6, we see the ECG wave from lead II of a normal sinus rhythm (NSR) i.e., the voltage between right arm and left leg. A classical ECG signal contains measurements from 12 such leads. One full cycle of ECG wave contains 6 peaks as shown in Figure 1.6, labeled as P, Q, R, S and T. Among all the peaks, the R peaks stand out the most in terms of amplitude and hence, is used as the standard reference to measure the cycle period. This is known as the RR interval and so the data used is called RR interval data. In terms of RR interval, normal resting heart rate is between 60 to 100 beats per minute (bpm) which translates to a duration of 0.6 to 1.2 seconds.

With this background, it is important to understand the relevance of measuring HRV in terms of RR interval data. Clinically, the correlation between reduced or increased HRV in a patient is well established to conditions such as myocardial infarction [Kleiger et al. 1987], congestive heart failure (CHF), depression etc., Hence, measurement and monitoring of RR intervals over extended periods of time in pa-

tients is a valuable diagnostic tool to identify the aforementioned ailments which can be fatal if undiagnosed.

The process of time keeping of heart beat is controlled by a group of cells located in the wall of right atrium of human heart known as pacemaker cells. These pacemakers cells are controlled by the sympathetic and parasympathetic nervous systems. Human nervous system is known to be divided into sympathetic, parasympathetic and enteric nervous systems each taking care of different aspects of human function. Sympathetic nervous system remains active at a basic level to initiate “flight or fight” response to maintain homeostasis or a state of constant equilibrium. The parasympathetic nervous system takes care of functions of internal organs that occur unconsciously and is responsible for “feed and breed” activities in the body. Enteric nervous system governs the function of gastrointestinal system and has its own reflex activity different from sympathetic and parasympathetic.

It is to be noted that the sympathetic and parasympathetic nervous systems act in opposing terms and are in a constant state of complex interaction to wrest control over the pacemaker cells [Levy 1971]. This design ensures that the human heart beat is maintained even when in an unconscious state such as the REM (rapid eye movement) sleep cycle. However, the process is known to be complex, seemingly nonlinear in nature [Brennan et al. 2001] due to complicated relationship between the two competing nervous systems. It is this underlying complexity that is of interest to our analysis.

In patients with NSR, it is known that these two systems are in near equilibrium, with some multi-scale nature. However, CHF patients i.e, those patients

whose heart is unable to pump increased volume of blood in time of increased body activity have a violation of homeostasis and that leads to a non-equilibrium between the two nervous systems. This implies an evident multi-scaled nature with processes of different complexities dominating over different time scales leading to violation of homeostasis. The analysis of crossover phenomenon is the objective of our work on the RR interval data.

## Financial data

Stock market index is a composite of several stock prices listed on the exchange. Dow Jones Industrial Average (DJIA), one of the oldest and longest such composite index comprising of 30 major American companies which was first calculated on May 26, 1896. A composite index is often estimated as weighted average of prices of the individual component stocks.

Financial markets and market movements are governed by a set of complex structures. They involve long term cycles or trends and speculative or regulatory short term effects. However, the complexity arises, among others, out of the involved human element which is based on sociopolitical and psychological trends [Cristelli 2014]. Such unpredictable drivers cause unpredictable outcomes and the evidence from past market booms and crashes supports this assertion e.g., boom of 1920, crash of 2008 etc., These fluctuations riding on top of known trends are the subject of much study and speculation in recent work.

Due to separation of time scales leading to the output time series of the index

values, financial markets are known to show multiple time scales. Particularly of recent interest is the study of nature of markets globally with respect to differing behaviors at short and long time scales within various indices. We direct our attention to apply our crossover analysis on financial index data and show the results in following chapters. Also, removal of predictable trends from market data yields the noise or the unpredictable component of the data. Our interest in characterizing predictability by studying this noise is another reason for choosing financial data as a data of interest in our studies.

### 1.3 Long Range Correlations and Hurst exponent

The phenomenon of long range correlations (LRC), sometimes referred to as long range dependency (LRD) or self similarity [Samorodnitsky and Taqqu 1994], involves study of decay in statistical correlations/dependence over time or spatial scales. Harold Edwin Hurst, a British hydrologist, was commissioned by the then government to study the long term storage capacity of reservoirs during the construction of Aswan low dam on the banks of river Niles, Egypt in the early 20<sup>th</sup> century. He was faced with the question of statistically measuring the increased levels of dam water during flood season and decreasing levels during droughts. His solution to measure such seasonality involved proposal of a metric, known as Hurst exponent ( $H$ ), which is a result of scale free nature of a measure of variability of changes in water levels [Hurst 1951] at various time windows. The now famous method measures this variability using a rudimentary technique known as rescaled



range analysis [Hurst et al. 1965].

Details of this method and an advanced scheme that forms the back bone of other schemes to measure the exponent in our work are shown in this section. LRC is explained in greater detail in following chapters in line with the context. Theoretical models espousing LRC phenomenon are shown in chapters of this thesis according to their application and relevance to the work.

## Rescaled Range analysis

Hurst wanted to measure the level of statistical correlations within a time series data, particularly stationary in nature. If the level of statistical dependence increases, then it was hypothesized that the variability in the time series data obtained as a random walk with the original data as the incremental process would increase. Hence, the variability in the random walk would yield clues to such statistical buildups in the incremental process. If  $X_1, X_2, \dots, X_n$  is a discrete stationary time series, then a mean adjusted series is constructed as  $Y_i = X_i - \sum_{j=1}^n X_j$ . This mean adjusting is to ensure that the data has a mean 0 similar to the usual gaussian noise  $\mathcal{N}(0, \sigma^2)$  with standard deviation  $\sigma$ .

From the mean adjusted time series  $Y_i$ , a cumulative deviate series or a random walk profile is constructed as  $Z_i = \sum_{k=1}^i Y_k$ . Clearly,  $Y_i$  is the incremental process of  $Z_i$  and hence, increased correlations in  $Y_i$  reflect in  $Z_i$  as its increased variability. This notion is central to any method of measuring LRC. Now, one needs to find the variability within  $Z_i$  and there is no unique method for this. Hurst proposed a metric

based on his observation that the range of  $Z_i$  within a walk size or time window length of estimation, normalized by its standard deviation within the window is known to be scale free in nature with respect to the window length. This observation is rooted in the fact that in a window of length  $t$ , range of a random walk scales as  $t^{0.5}$  and if there were no correlations then the standard deviation would not depend on  $t$ , i.e., the rescaled range would scale as  $t^{0.5}$  in an uncorrelated data, yielding an exponent of  $H = 0.5$ .

Mathematically, the range  $R_t$  within a time window  $t$  is constructed as  $R_t = \max\{Z_1, Z_2, \dots, Z_t\} - \min\{Z_1, Z_2, \dots, Z_t\}$  and  $S_t$  is standard deviation of the series  $\{X_1, X_2, \dots, X_t\}$ . The rescaled range is estimated as  $R_t/S_t$  and it scales as  $t^H$ . So, a linear regression between logarithms of  $R_t/S_t$  and  $t$  would yield a slope  $H$ .

As mentioned earlier, Hurst envisioned his metric for a stationary time series such as gaussian noise. In fact, a well known theoretical model with normal distribution that demonstrates LRC called fractional Gaussian noise (fGn) and its random walk equivalent fractional Brownian noise (fBm) was proposed [Mandelbrot and Van Ness 1968]. However, it was found that the rescaled range analysis had lot of loop holes due to the primitive way of estimation of variability. To increase the robustness of estimation of  $H$ , an alternative method of measuring variability was proposed [Karlin and Brendel 1993] known as fluctuation analysis.

## Fluctuation Analysis

Fluctuation analysis (FA) shares the same basis of rescaled range analysis up to the point of constructing the random walk profile  $\{Z_i\}$  of a time series  $\{X_i\}$ . The major deviation of this scheme from rescaled range is in the way the variability of  $\{Z_i\}$  is measured. Fluctuation functions are estimation as a function of window size as  $F(t) = \langle (Z_{i+t} - Z_i)^2 \rangle$ , where  $\langle \dots \rangle$  refers to the ensemble average for all realizations starting at various initial times  $i$  within the data.

This fluctuation function  $F(t)$  is extensively used over the course of our work. Crossovers were found to exist in the logarithmic plot of  $F(t)$  versus time window  $t$  for multi-scale data. In data without multiple time scales or negligible crossover phenomenon, the fluctuation functions demonstrate a scale free nature just like the rescaled range. Estimation of Hurst exponent  $H$  is performed using the same approach of linear regression between logarithms of  $F(t)$  and  $t$  as mentioned earlier.

It was found that both rescaled range and fluctuation analysis suffer from a common limitation—they fail to estimate an exponent in non-stationarity data. As there is no common approach to detrend a time series of trends *a priori*, an alternative that proposes to detrend polynomial trends of fixed order was proposed, known as Detrended Fluctuation Analysis (DFA) [Peng et al. 1992] and described in more detail in Chapter 2.

## 1.4 Crossover phenomenon in fluctuation functions and probability distributions

DFA technique is extensively applied to analyze correlations within non-stationary time series data. It assumes that the random walk profile  $\{Z_i\}$  can be separated into two independent parts i.e., the polynomial trend  $T$  and the signal  $S$  as:  $Z_i = T_i + S_i$  with null covariance structure i.e.,  $\text{Cov}(T, S) = 0$ . This assumption is not always true and is unsupported from the outcome of DFA application on several real world data.

In fact, data corresponding to many records show that crossovers exist within the fluctuation functions in FA or DFA [Hu et al. 2001, Liu et al. 1999, Movahed et al. 2006]. This supports our assertion that multi-scale nature of data manifests itself as crossover time scales separating regimes in DFA or FA that use fluctuation functions. We propose to address such crossovers using a new technique as described in Chapter 3. Results from its application to space weather, heart rate variability and financial index data are shown in Chapters 3, 4 and 5 respectively.

Crossovers are also known to exist in probability distributions of several complex data. In fact, these crossovers help explain heavy tailed distributions. In Figure 1.7, we see the evidence of heavy tail nature in the probability distribution of AL index data when compared to normal distribution drawn using mean of AL as its mean and median as its variance. This data consists of AL index values for the past 33 years from 1980-2013 with over 14 million data points and hence can

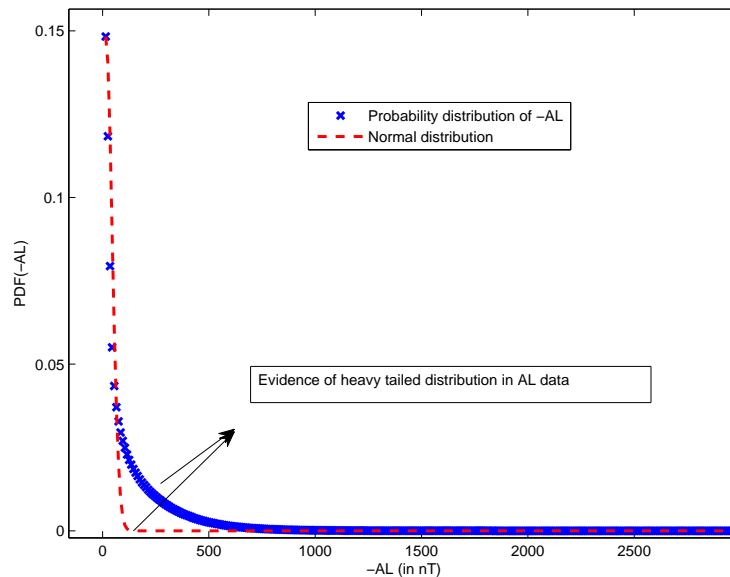


Figure 1.7: Probability distribution of AL index data from 1980-2013 (in blue). Comparison between observed probability and normal distribution (in red) shows the evidence of heavy tailed nature of distribution.

be considered comprehensive enough to model heavy tail in the distribution. Heavy tailed distribution in a probability distribution  $P(x)$  can be described mathematically as:  $P(x > x_u) \propto x^{-\mu}$  for events,  $x \geq x_u$ , where  $x_u$  is a cutoff magnitude above which heavy tails exist in the distribution. The exponent,  $\mu$  is commonly referred to as the tail exponent and can be estimated using a variety of estimator techniques.

Traditionally risk in finance was modeled as a normal distribution without heavy tails as events whose magnitudes deviate more than 5 times the standard deviation (known as 5 sigma events) from the mean of data are usually rarely found. Existence of heavy tails leading to undefined variance complicate predictability of financial data violating normal distribution model for risk. This was identified earlier [Mandelbrot 1963; 1997, Taleb 2007] and it was found that a more general stable

distributions better model returns in finance as against normal distributions. The extreme events are usually called black swans due to the rarity of their occurrence in nature compared to abundantly seen white swans.

Due to the existence of crossovers both in fluctuation functions and probability distributions, its analysis is a crucial step in identification of scales, tail exponents in data and hence motivated our work in this direction.

## 1.5 Nonlinear time series prediction

Time series forecasting involves modeling of historical data using the resulting model to predict/forecast the values at future time step(s). Statistical methods prescribe to model past data using parametric or non-parametric regression schemes. These regression models are based on assumptions about the nature of the data ahead of time, often introducing artefacts due to its presumptive aspect.

Dynamical systems theory proposes an alternative way of modeling historical data. Under this scheme, based on Taken's theorem [Takens 1981], the phase space corresponding to states of the system and the evolution of their trajectories can be reconstructed using a univariate time series of the system. It must be noted that this approach is a semi supervised machine learning scheme as the phase space reconstruction is data derived and involves few to none modeling parameters.

As an example, starting from the past data of  $x(t)$  of the three dimensional Lorenz attractor, one can reconstruct the phase space whose  $y(t), z(t)$  were unknown to begin with. Using the reconstructed phase space, one can retrieve information

about possible evolution paths from the most recent state. These predicted evolutions can be averaged using a preselected method such as a weighted or unweighted mean for what are the closest states i.e., nearest neighbors in the reconstructed phase space [Chen and Sharma 2006, Ukhorskiy et al. 2002].

This method not only affords predictive value as an alternative to the usual machine learning schemes based on statistical or probabilistic inference approaches, but also yield us an estimate of trends in data. Using this technique, we uncover the underlying dynamical features in the data. Trends in data are due to seasonalities and other long term variations in the data. These correlated effects reflect in data as dynamics. Due to this equivalence of trends and dynamics, we propose to use nonlinear time series predictions as a measure of trend in system to be removed to yield the remnant fluctuations or noise (short term component) within the data. Hence, understanding this technique is a necessary precursor to follow our work and analysis in chapter 6 of the thesis.

Time delay embedding and phase space reconstruction are the key elements in data derived modeling to achieve time series forecasting based on mean-field or local-linear filters. Finding the appropriate time delay and the embedding dimension are the primary steps to reconstruct the phase space. We show some of the more common techniques used to find the time delay, embedding dimension and finally demonstrate how this reconstructed phase space can be used for prediction when applied to Lorenz attractor.

## Lorenz attractor-the model

We demonstrate the time series prediction technique using time series data generated from the well known Lorenz attractor [Lorenz 1963], which is a three dimensional, nonlinear, deterministic system:

$$\begin{aligned}\frac{dx(t)}{dt} &= \sigma(y(t) - x(t)), \\ \frac{dy(t)}{dt} &= -x(t)z(t) + rx(t) - y(t), \\ \frac{dz(t)}{dt} &= x(t)y(t) - bz(t).\end{aligned}\tag{1.2}$$

The above equations are derived by finite mode truncation of the partial differential equations describing the thermal driving of convection in the lower atmosphere. The parameters used in solving the system of equations in equation 1.2 are  $r = 28$ ,  $b = 8/3$ , and  $\sigma = 10.0$  with an integration time step of  $dT = 0.01$ .

## Phase space reconstruction

An inherent property of a nonlinear dissipative system i.e., contraction of phase space as the system approaches its asymptotic state allows us to model the chaotic system by a few nonlinearly coupled ordinary differential equations. Such a nonlinear system has a characteristic attractor which is a set towards which a variable, moving according to the equations of the dynamical system and which evolves over a span of time. Its resulting attractor is characterized by a dimension, a minimum number of independent variables required to obtain the complete information about the attractor. A complete picture of this phase space can be obtained from one or more of the system variables given that they are sufficiently nonlinearly coupled [Packard



et al. 1980, Ruelle 1989, Takens 1981]. Time delay embedding is a technique used to reconstruct this phase space from a time series data. In this technique, we begin by splitting our time series into  $m$ -component phase vectors  $X_i$  such that,

$$X_i = \{x_1(t_i), x_2(t_i), \dots, x_m(t_i)\} \quad (1.3)$$

where  $x_m(t_i) = x(t_i + (m - 1)T)$ ,  $T$  is an appropriate time delay.

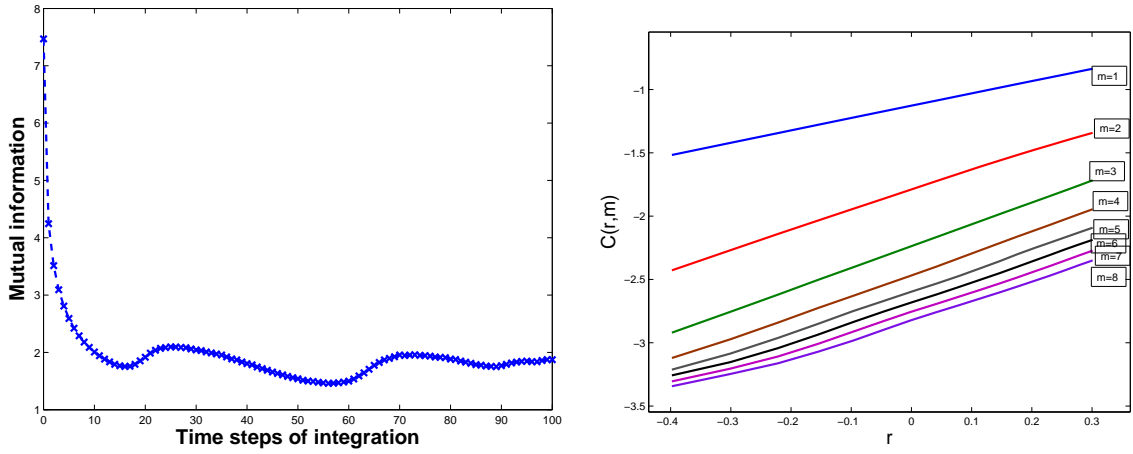
The process of choosing an appropriate dimension  $m$  is known as embedding. We need a dimension  $m \geq 2m_A$ , the attractor dimension, so as to obtain an attractor in the reconstructed phase space which is smoothly related to the attractor in the original phase space. We now illustrate the most common techniques used for choosing the time delay and embedding dimension for Time delay embedding method [Abarbanel et al. 1993].

## Mutual Information

Calculating mutual information function between two sets of variables helps us understand the extent of correlations between the two variables [Shannon 1948]. Mathematically, if  $x_1$  and  $x_2$  are the two random system variables with  $i_1$  and  $i_2$  as the possible outcomes of the respective system measurements, then the average mutual information between the measurements is defined as [Cover and Thomas 1991]:

$$I(x_1, x_2) = \sum_{i_1} \sum_{i_2} p_{12}(i_1, i_2) \log_2 \left[ \frac{p_{12}(i_1, i_2)}{p_1(i_1)p_2(i_2)} \right] \quad (1.4)$$

where  $p_1(i_1)$  and  $p_2(i_2)$  are the probability distributions of the random variables  $x_1$  and  $x_2$  assuming values  $i_1$  and  $i_2$  respectively and  $p_{12}(i_1, i_2)$  is their joint probability



(a) Mutual information of Lorenz attractor. (b) Correlation integrals analysis of Lorenz.

Figure 1.8: Estimation of time delay and embedding dimension parameters for phase space reconstruction of Lorenz attractor.

distribution.

We observe from equation 1.4 that when  $x_1$  and  $x_2$  are independent,  $I(X_1, X_2) = 0$  since  $p_{12}(i_1, i_2) = p_1(i_1)p_2(i_2)$ . This average mutual information function includes the higher order correlations than the linear correlation function thus yielding a better measure for nonlinear systems. We can extend this definition of mutual information to the context of a single time series,  $x(n)$  in order to calculate the appropriate time delay or a good measure of the characteristic time of the system. The mutual information in equation 1.4 can be calculated in the case of time series  $x(n)$  as a function of time delay  $T$  between  $x(n)$  and  $x(n + T)$  [Gallager 1968]:

$$I(T) = \sum_{n=1}^N p(x(n), x(n + T)) \log_2 \left[ \frac{p(x(n), x(n + T))}{p(x(n))p(x(n + T))} \right] \quad (1.5)$$

By calculating  $I(T)$  from equation 1.5, we can find the appropriate time delay  $T_d$  as where the first minimum of  $I(T)$  occurs [Fraser and Swinney 1986]. This

minimum of  $I(T)$  for the Lorenz attractor can be seen in Figure 1.8a, where the first minimum of  $I(T)$  occurs when time delay  $T_d = 16$  steps of integration which corresponds to 0.16 seconds of time for our Lorenz attractor and this value would be used as the time delay for phase space reconstruction.

## Correlation integral

With the time delay,  $T_d$  obtained from Mutual information analysis, we can find the embedding dimension  $m$  by calculating a sequence of Correlation integrals. This process relies on the principle that once a desired minimum embedding dimension is achieved, the properties of the resulting attractor which depend on the distance between points in the respective phase space should be invariant with increasing embedding dimension. One such property we use is Correlation integral, which is defined for  $N$  vectors in an  $m$ -dimensional space for the distance  $r$  as [Abarbanel et al. 1993]:

$$C(r, m) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \Theta(r - |X_i - X_j|) \quad (1.6)$$

where  $\Theta$  is the Heavyside function.

The correlation dimension can then be defined as  $\nu = \lim_{r \rightarrow 0} \frac{\log(C(r, m))}{\log(r)}$ . From Figure 1.8b, we see that the slope of this correlation integral is invariant for  $m \geq 3$ . So, for phase space reconstruction, we choose  $m = 3$  as the embedding dimension.

## Time series prediction

The reconstructed phase space can be used to predict possible future time steps from the most recent state. We first define a trajectory matrix as an  $N \times m$  matrix which is a collection of  $m$  dimensional state vectors  $X_i$  obtained from time delay embedding. That is, a trajectory matrix  $X$  can be written as:

$$X = N^{-1/2} \begin{bmatrix} x_1(t_1) & x_2(t_1) & \dots & x_m(t_1) \\ x_1(t_2) & x_2(t_2) & \dots & x_m(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_N) & x_2(t_N) & \dots & x_m(t_N) \end{bmatrix} \quad (1.7)$$

where  $N$  is the number of vectors. This  $N \times m$  matrix contains the dynamical features of the system contained in the time series data and can be used to find the evolution in the reconstructed phase space.

Mean Field Filter technique [Ukhorskiy et al. 2002] is commonly used to predict future state i.e.,  $X_{n+1,pred}$  is obtained as an average of evolved states of all the nearest neighbors ( $NN$  in number) of the  $n^{th}$  state  $X_n$ . A nearest neighbor is defined as a state closest to a state of interest in terms of the euclidean distance. The average value of the state vectors of the nearest neighbors is defined usually as the center of mass.

$$X_{n+1,pred} = \frac{1}{NN} \sum_{k=1}^{NN} X_{k,n} \quad (1.8)$$

where  $\{X_{k,n}\}$  are the evolved states of the nearest neighbors of the most recent state  $X_n$ .

Weighted Mean Field Filter [Chen and Sharma 2006] is another technique

essentially same as Mean Field Filter except that we attach a set of weight factors  $g$  which depend inversely on the distances of the nearest neighbor from the center of mass. This should improve the accuracy of prediction because not all nearest neighbors are identically spaced from the center of mass.

$$g_k = \frac{1}{d_k^2} / \sum_{i=1}^{NN} \frac{1}{d_i^2} \quad (1.9)$$

where  $d_i$  is the Euclidean distance of the  $i^{th}$  nearest neighbor from the center of mass. The resulting predicted future state i.e.,  $X_{n+1,pred}$  is calculated as:

$$X_{n+1,pred} = \frac{1}{NN} \sum_{k=1}^{NN} X_{k,n} \bullet g_k \quad (1.10)$$

In these local linear filters, we use the linear term in an expansion around the initial conditions and since the linearity of expansion is local, we conserve the nonlinearity of the system as in the reconstructed phase space. To quantize the error in the predictions, Normalized Mean Square Error (NMSE) is routinely used:

$$\eta = \frac{1}{\sigma_0} \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - X_{i,pred})^2} \quad (1.11)$$

where  $X_i$  and  $X_{i,pred}$  are the observed and predicted data respectively and  $\sigma_0$  is the standard deviation of actual data  $\{X_i\}$ .

Using the previously defined prediction techniques, we predict the future states of the time series corresponding to  $x$  component of Lorenz attractor and quantify the accuracy of prediction with NMSE measurements. As previously shown, values of time delay,  $T_d = 16$  time steps and embedding dimension,  $m = 3$  were used in the phase space reconstruction. The number of nearest neighbors,  $NN$  were taken to be 20 for our predictions. Resulting predictions utilizing weighted mean field filter

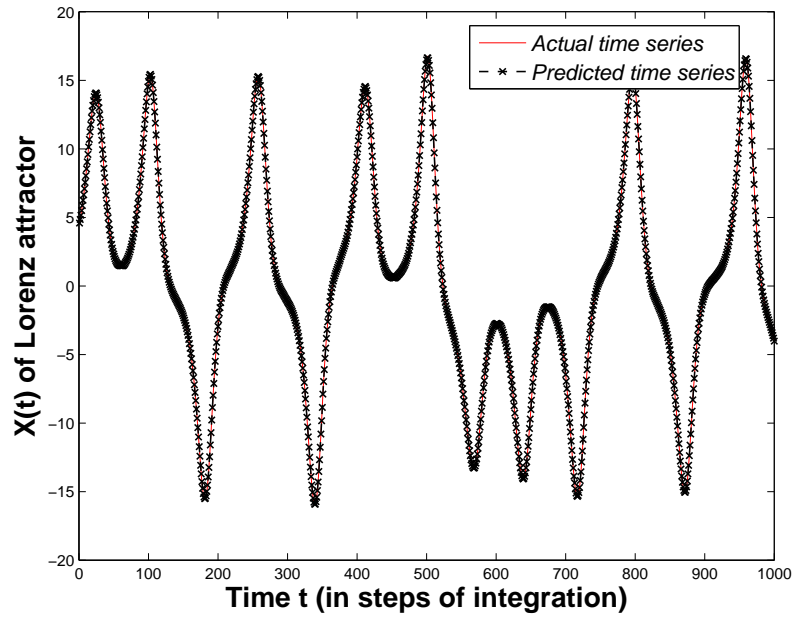


Figure 1.9: Prediction of  $x(t)$  of Lorenz attractor using weighted mean field filter technique.

technique are shown in Figure 1.9. The NMSE values for the weighted mean field filter technique as applied to predict  $x(t)$  of the Lorenz system is calculated to be 0.0046.

The data and techniques that would be used in this work are presented so far. We present a brief outline of the contents of the thesis to conclude this introduction.

## 1.6 Outline of Thesis

In Chap. 2, DFA is introduced and its application demonstrated when applied to a model with a time varying LRC - multifractional Brownian motion (mBm). This serves as an introduction to DFA and also serves to benchmark its performance. Also, evidence of mBm-like nature in AL index data is shown.

In Chap. 3, crossover analysis in fluctuation functions of AL index data is presented. Technique for such an analysis is introduced and subsequent results follow. A suitable model for crossover phenomenon is presented using which salient features such as crossover time in AL data are robustly estimated.

In Chap. 4, crossover features in physiological data are shown. Analysis of the extent of crossover helps us propose a metric for diagnostic purposes to identify ailments such as congestive heart failure.

In Chap. 5, crossovers in fluctuation analysis of generalized volatility of financial indices are shown. Analysis of crossover in terms of increasing LRC versus decreasing LRC is presented.

In Chap. 6, Fluctuation Analysis after Trend Elimination (FATE) is demonstrated on a theoretical model with separable trend and signal. FATE on data from finance and AL index is presented. Correlation is made between Hurst exponent from FATE and tail exponent in the probability distribution of noise which serves as an innovative method to characterize predictability.

In Chap. 7, conclusions are given with respect to work from each chapter and on an ending note possible future work is discussed.

## Chapter 2: Characterizing Detrended Fluctuation Analysis of Multifractal Brownian Motion

### 2.1 Overview

We introduced the concepts behind long range correlations and multi-scale systems in the previous chapter. Detrended Fluctuation Analysis (DFA) is an important statistical analysis tool to estimate scaling in systems with long range correlations. Multi-scale systems are defined to be those complex systems whose component processes are of varied correlation time scales. Separability of these time scales via scaling crossover in fluctuation analysis is proposed to be the basis for study of these component processes.

Multifractal systems are also multi-scale in nature as they incorporate processes of varying fractal dimensions. Multifractal Brownian motion is a special class of multifractal systems where the fractal dimension of the Brownian motion is time varying through the time span of the data. In this chapter, we apply the popular DFA technique towards artificial multifractal data as a first step in testing its robustness when applied to data with such a generic multifractal nature. Also, this chapter serves as an introduction to important features of DFA technique which



would come up again in latter chapters of this thesis as a result of its widespread application.

Statistical properties such as trends and correlations of complex phenomena are important in the study of the mechanisms driving such systems to extreme events. Due to the non-equilibrium nature of complex driven systems, general statistical analysis tools cannot be readily applied on them. Long Range Correlations (LRC) is a key feature and is studied in data from diverse physical systems such as temperature records [Koscielny-Bunde et al. 1998], river flows [Mandelbrot and Wallis 1969], heart rate variability [Schäfer et al. 1998], space weather [Sharma and Veeramani 2011] etc.,

Rescaled range analysis (R/S) [Hurst 1951] and fluctuation analysis (FA) [Karlin and Brendel 1993] are statistical tools developed to estimate the variability of time series through estimation of Hurst exponent,  $H$  [Hurst et al. 1965], a statistic which is directly related to the scaling in auto-correlation functions, and, also to the fractal dimension of the time series data. While the scaling exponent,  $H$ , is equal to 0.5 for uncorrelated white noise, many natural systems demonstrate values close to 0.7 [Sutcliffe 1979]. An introduction to these concepts was provided in Section 1.3. These techniques fail to estimate  $H$  in non-stationary data and this is their main shortcoming.

More recently, DFA was developed [Peng et al. 1994] which is widely considered a better technique than R/S or FA due to its capability to detrend a time series data whilst estimating  $H$ , making it viable for non-stationary systems. With increased usage of DFA technique, scepticism regarding its detrending capabilities

arose [Bryce and Sprague 2012] and it is evident that there is need for a better alternative detrending schemes for data with atypical trends (e.g., nonlinear trends). In spite of its purported shortcomings, DFA is recognized as an efficient Hurst exponent estimation technique. Its efficiency stems from the fact that due to simultaneous detrending DFA estimates over lesser number of averages than in FA and thus faster.

Fractional Brownian motion (fBm), a generalization of Brownian motion is a quintessential representative model of the Hurst effect [Mandelbrot and Van Ness 1968]. Since its discovery, there has been increased interest in modeling physical systems as fBm. However, it was quickly realized that imposing a uniform  $H$  over the span of the data is in fact a restricting condition as uniform level of LRC in real life data implies mono-fractal nature. This is a very strong condition and hence, is not commonly found in real world data. Multifractional Brownian motion (mBm) is a generalization of fBm relaxing this condition [Peltier and Lévy-Véhel 1995], allowing for variable degrees of self-similarity with non-stationary increments i.e.,  $H$  varies as  $H(t)$  over the time span of the data. Control over  $H(t)$ , which is a measure of LRC at each point of time  $t$ , is a valuable feature of mBm due to which there has been increased interest in modeling various geophysical systems as mBm [Echelard et al. 2010, Gaci and Zaourar 2011, Wanliss and Dobias 2007].

Although there is increasing use of DFA as a technique to study LRC in time series data, it is widely recognized that it yields a single Hurst exponent, and thus can not distinguish between multi-fractal and mono-fractal systems, e.g., between mBm and fBm respectively. In fact, most systems exhibit time varying  $H$  exponent, but DFA yields a constant  $H$  value because it is a single exponent

technique. Besides which previous studies have shown the effect of data size used on the Hurst exponent [Lennartz and Bunde 2011, Rybski and Bunde 2009], thus requiring caution in the interpretation of the estimated values. This is in direct agreement with our study of effect of data size on the Hurst exponent estimated by DFA in mBm data as seen in Section 2.5.

Although other schemes such as Multi Fractal Detrended Fluctuation Analysis (MF-DFA) were proposed [Kantelhardt et al. 2002] as an alternative, they address the multifractal nature of time series with respect to one fractal dimension at a time and do not provide a solution with respect to estimating the time varying fractal structure of mBm. It is apparent that DFA and other similar techniques were assumed to locally estimate a time averaged Hurst exponent [Carbone et al. 2004, Sheng et al. 2010]. This assumption helps compute  $H(t)$  by using estimator techniques with sliding windows. We believe the success of estimating such a time average depends on the assumption of local linearity of the Hurst exponents. This is an assumed result with no systematic study performed in the past and brought to fore the question- how does DFA perform in estimating a time average exponent in mBm data?

Using mBm data generated from linearly varying Hurst exponents,  $H(t)$ , we show that DFA in fact estimates the time average (of  $H(t)$ ) and test the dependence of estimated exponent on various data and technique related parameters. The primary motivation for our study is to establish a bench mark for the performance of DFA technique with respect to its ability to estimate a time averaged Hurst exponents from mBm data, and, identify its limits.

This work characterizes the performance of DFA on mBm data with linearly varying  $H(t)$  and further test the robustness of estimated time average with respect to various data and technique related parameters. Our results agree with what is intuitively expected from such a study and set to serve as a bench-mark for using DFA as a sliding window estimator to obtain  $H(t)$  in a time series data. Furthermore, given that DFA is one of the primary techniques used for fluctuation analysis throughout the work mentioned in the current thesis, this chapter serves to set the context and understand the limits and limitations of applicability of DFA to data with time varying fractality.

The next section introduces preliminary ideas of LRC, fBm, mBm, and, DFA. We then proceed to establish our estimation methodology and follow it up with main results of this study and conclusions.

## 2.2 Fractional and Multifractional Brownian Motion

LRC is also commonly identified as self-affinity, self-similarity or long-range persistence and is a statistical property of time series where the rate of decay of its auto-correlation is slower than exponential decay and most commonly a power law. This property is usually quantified using the Hurst exponent  $H$  (also known as the Hölder exponent). The Hurst exponent,  $H \in (0, 1)$  with increasing value of exponent implying increasing LRC nature of the data, and 0.5 as threshold value where the correlations are completely absent.

Fractional Gaussian noise (fGn, which is stationary in nature) is proposed

as a model for data with LRC and fractional Brownian motion (fBm) (its non-stationary counterpart) is its corresponding Wiener process generated using fGn as its incremental process. A continuous time fractional Brownian motion (fBm),  $B_H(t)$  with Hurst exponent  $H$  is a Gaussian process with zero-mean and is  $H$ -self affine i.e.,

$$B_H(\lambda t) \cong \lambda^H B_H(t), \forall \lambda > 0 \quad (2.1)$$

Also, its covariance varies by definition as [Kamont 1996],

$$cov[B_H(t_1), B_H(t_2)] = \frac{1}{2}(|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}) \quad (2.2)$$

Thus,  $H$  characterizes the relative smoothness of such resulting Brownian motions. It can also be seen that when  $H = 1/2$  and  $t_1 > t_2$  (without loss of generality),  $cov[B_H(t_1), B_H(t_2)] = t_2$  (from Equation 2.2), thus it is a Wiener process (Brownian motion). However when  $H > 1/2$ , then the increments are positively correlated and when  $H < 1/2$ , the increments are negatively correlated. This means that we have a smooth long-term correlated time series data for when  $H > 1/2$  and anti-correlated data for when  $H < 1/2$ .

$$G_H(t) = B_H(t + 1) - B_H(t) \quad (2.3)$$

The increments in a fBm is known as fractional Gaussian noise (fGn) and is seen in Equation 2.3 as  $G_H$ . It is stationary in nature. One obvious problem with fBms is that although they capture the self-similarities well, the point wise irregularity given by the constant Hurst parameter,  $H$  is invariant in time. This restricting condition can be over come by generalizing fBm to a more broader class of Brownian motions

where the local irregularity is given by a time varying Hurst exponent,  $H(t)$ , known as multifractional Brownian motion (mBm).

In order to simulate data with such properties, an fBm  $B_{H(t)}$  with a Hurst parameter,  $H(t)$  at a time  $t$  is generated and a mBm,  $W_{H(t)}$ , is calculated to be equal to  $B_{H(t)}$  by interpolation. Each of a field of such fBms can be generated using many methods. In our simulations, we applied the method of circulant matrices [Wood and Chan 1994] to generate each of such fBm with different Hurst exponents. Producing an fBm with a stationary fGn as its incremental process (as in Equation 2.3) requires a method to filter a typical Gaussian noise of low frequency components to produce a covariance as described in Equation 2.2 using Fourier filtering technique. Circulant matrices, which are diagonalized by a discrete Fourier transform allow for faster solutions for linear equations that contain them [Davis 1970], are used to obtain each fBm. Subsequently, the field of fBms are interpolated by a kriging method [Chan and Wood 1998]. The method of kriging predicts intermediate value of a function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point, as against a piecewise-polynomial spline, which based on smoothness may not produce the most likely values for intermediates. This method is known to be reliable in simulating discrete time mBm [Coeurjolly 2000]

### 2.3 Detrended Fluctuation Analysis

Detrended Fluctuation Analysis (DFA) is a widely used technique for quantifying LRC in time series data, particularly those with non-stationarities [Kantelhardt

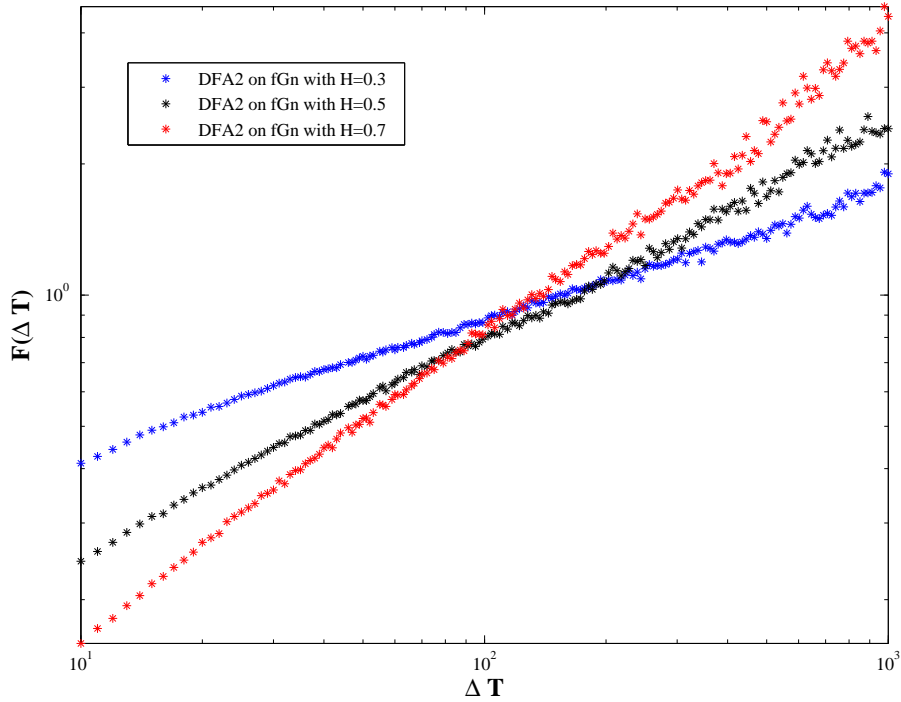


Figure 2.1: Application of Detrended Fluctuation Analysis with quadratic detrending (DFA2) to fractional Gaussian noise with Hurst exponents,  $H=0.3$ ,  $0.5$  &  $0.7$ .

et al. 2001]. In this technique, a random walk profile  $Y(j)$  is created using the time series as is usually done in FA (shown in Section 1.3). However, its departure from traditional FA arises due to the simultaneous detrending operation performed using a polynomial of known order ( $p$ ) which is locally fit to the walk within a time window of fixed size. DFA, thus consists of the following steps: The time series data  $X(i)$  of length  $N$  is first shifted by its mean  $\langle X \rangle$  and its cumulative sum calculated as:

$$Y(j) = \sum_{i=1}^j [X(i) - \langle X \rangle] \quad (2.4)$$

This cumulative sum series,  $Y(j)$  is now segmented into time windows of different lengths  $\Delta T$  thus yielding a collection of set of random walk profiles of varying sizes. These profiles are detrended within the windows by estimating their trends as best fit polynomials of order  $p$ ,  $Y_{\Delta T}^{(p)}(j)$  within each box corresponding to size  $\Delta T$ . These trends are now removed by subtracting from the cumulative sum data and the resulting fluctuations, generally referred to as fluctuation functions are calculated as the mean squared deviations i.e.,

$$F^{(p)}(\Delta T) = \left( \frac{1}{N} \sum_{j=1}^N [Y(j) - Y_{\Delta T}^{(p)}(j)]^2 \right)^{\frac{1}{2}} \quad (2.5)$$

If the data  $X(i)$  have long-range correlations, it would be reflected in the fluctuation functions  $F^{(p)}(\Delta T)$ , with a power-law as:

$$F^{(p)}(\Delta T) \propto (\Delta T)^\alpha \quad (2.6)$$

The scaling exponent  $\alpha$  is calculated from scaling within DFA by best fit of log-log plot between  $F^{(p)}(\Delta T)$  versus window size  $\Delta T$  using linear regression (method of least squares in our work). This scaling exponent,  $\alpha$  is related to the Hurst exponent  $H$  as  $\alpha = H$  for fGn and as,  $\alpha = H + 1$  for fBm because fBm is the cumulative sum or the integral of fGn as seen in Equation 2.3. It must be noted that the Hurst exponent for both fGn and fBm is equal to 0.5 because it is a measure of LRC in the corresponding systems, which are of the same extent given that one is an incremental process of the other. However, as  $H$  is not directly measurable in fBm due to its non-stationarity,  $\alpha$  is used as a measure of  $H$ .

It is very important to note that this exponent  $\alpha$  can be related to other well known exponents such as  $\beta$  in power spectrum decay ( $P(f) \propto f^{-\beta}$ ) and  $\gamma$  in the



power law decay in correlation function ( $C(L) \propto L^{-\gamma}$ ) as:

$$\begin{aligned}\alpha &= 1 - \gamma/2, \& \\ \alpha &= (\beta + 1)/2\end{aligned}\tag{2.7}$$

The first relationship between  $\alpha$  and  $\gamma$  is due to the dependence of fluctuation functions in DFA on correlation functions [Kantelhardt et al. 2001]. The second relationship between  $\alpha$  and  $\beta$  is due to the fact that correlation functions are the fourier transforms of the spectral functions in power spectrum [Buldyrev et al. 1995]. Analogous relationships between  $H$ ,  $\beta$  and  $\gamma$  can be drawn given the stationary or non-stationary nature of the data.

DFA technique applied with a detrending of  $p^{th}$  order polynomial is commonly referred to as DFA<sub>p</sub>. Figure 2.1 shows the application of DFA2 to fGn data with various  $H$  exponents. The fGn data was produced using a popular wavelet-based synthesis method [Abry and Sellan 1996] details of which are irrelevant to the current work. From the results in Figure 2.1, we see that  $H$  is the slope of fluctuation functions from DFA2. Hence, this shows the application of DFA technique to measure  $H$  exponent, a measure of LRC in time series data.

We now characterize the performance of DFA on systems with time varying  $H$  exponents using simulated mBm data with linearly varying  $H(t)$  in time. The choice of linear  $H(t)$  is motivated by the consideration that its time variation will be gradual, and, the system can be treated as piece-wise local linear (in Hurst exponents). This allows one to divide the data into segments in which a time averaged Hurst exponent can be estimated under the local linearity assumption. It should be noted that piece-wise linearity forms the basis for earlier studies [Sheng

et al. 2010]. The results of the time averaged Hurst exponent estimated by DFA technique were found to be in accordance with results from linearly varying  $H(t)$ . That is, robustness of  $H$  estimated by DFA for mBm data with a sinusoidally varying  $H(t)$  was found to depend on the amplitude and the frequency (of  $H(t)$ ). More specifically, increasing the amplitude or frequency of the sinusoidal variation was found to deteriorate the quality of  $H$  estimated by DFA. This can be explained by the effect of slope of linear  $H(t)$  on the quality of exponent estimated by DFA as seen in Section 2.5. Hence, results from linearly varying Hurst exponents can be used to explain results for other form of  $H(t)$  in real life data by using local linearity arguments.

## 2.4 Estimation of time averaged Hurst exponent by DFA

The discrete time mBm data is simulated with  $H(t) = at + b$  over a time length  $t = 1$  to  $T$  using the data simulation method described in Section 2.2. The parameters  $a$  and  $b$  are chosen to control the range of  $H(t)$  i.e.,  $R = \max_{1 \leq t \leq T} H(t) - \min_{1 \leq t \leq T} H(t)$ , and, its mean =  $\langle H \rangle_{act}$ . This condition relates various parameters as:

$$\begin{aligned} a &= \frac{R}{T-1}, \& \\ b &= \langle H \rangle_{act} - \frac{R}{2} - \frac{R}{T-1} \end{aligned} \tag{2.8}$$

The time series of the mBm generated using technique described in Section 2.2 is shown in Figure 2.2.

When  $H_1(t) = 0.75$ , the mGn corresponding to respective mBm demonstrates uniform variability as expected. For  $H_2(t) = at + b$  with  $a = 10^{-5}$  and  $b = 0.7$

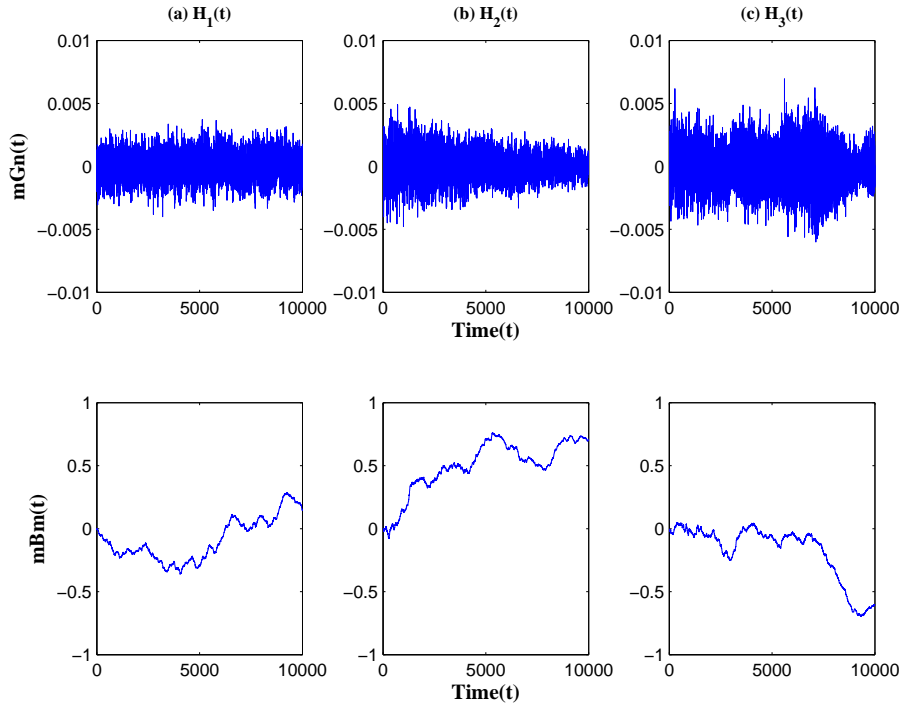


Figure 2.2: Simulated mBm data with various forms of  $H(t)$  with the incremental process, mGn shown in upper panel and the corresponding mBm in bottom panel. In (a),  $H_1(t) = c = 0.75$ . In (b),  $H_2(t) = at + b$  with  $a = 10^{-5}$  and  $b = 0.7$  i.e.,  $\langle H \rangle_{act} = 0.75$  and  $R = 0.1$ . In (c),  $H_3(t) = H_2(t) + \epsilon_t$  with  $\epsilon_t$  drawn from normal distribution,  $\mathcal{N}(0, \sigma^2)$  such that  $\sigma = \langle H \rangle_{act} \times 10^{-3}$ . It is interesting to note the differences in variability in mGn resulting in differences in LRC natures of mBm.

i.e.,  $\langle H \rangle_{act} = 0.75$  and  $R = 0.1$ , we see that with increasing  $t$ ,  $H$  increases resulting in decreasing variance in mGn data thus resulting in increasing LRC in mBm data when compared to the case in Figure 2.2(a). In Figure 2.2(c), for  $H_3(t) = H_2(t) + \epsilon_t$  with  $\epsilon_t$  drawn from normal distribution,  $\mathcal{N}(0, \sigma^2)$  such that  $\sigma = \langle H \rangle_{act} \times 10^{-3}$ , we see that due to noise in  $H_3(t)$  data, the variance shows no regular pattern in mGn and that is reflected in the LRC pattern in mBm. These figures show us qualitatively the difference in the nature of LRC within data with different forms of

$H(t)$ . If we had no *a priori* knowledge of the time varying nature of a data's Hurst exponent, the differing degrees of LRC between data would be viewed assuming a uniform exponent. Similarly, a single Hurst exponent estimation technique such as DFA would measure the time averaged exponent, denoted as  $\langle H \rangle_{est}$  which would be equal to  $\langle H \rangle_{act}$  under the ideal condition of  $R = 0$  (or  $a = 0$ ).

While measuring for  $\langle H \rangle_{est}$  using DFA, we apply linear regression between logarithm of fluctuation function ( $F(\Delta T)$ ) and logarithm of the time window ( $\Delta T$ ), specifically the method of ordinary least squares to find the best-fit slope and the standard error of least squares slope [McCuen 1985]. We measure the slope (or Hurst exponent) resulting from the linear regression over logarithmically equi-spaced values of time windows, ranging from  $\Delta T = 10$  to  $\lfloor \frac{T}{10} \rfloor$  (the integer rounded floor value) to eliminate spurious curvatures at lower time windows and also have certain minimum number of samples to obtain the relevant averaged fluctuation function at higher time windows.

We denote and measure the standard error as  $SE(\langle H \rangle_{est})$  (shown in error bars in the following figures) which gives us the precision in  $\langle H \rangle_{est}$  estimated by the DFA technique. Furthermore, error of estimation of  $\langle H \rangle_{act}$  is measured in terms of relative error of estimation as:

$$EE(\langle H \rangle_{est}) = \frac{|\langle H \rangle_{est} - \langle H \rangle_{act}|}{\langle H \rangle_{act}} \quad (2.9)$$

For some of the sections, we perform ensemble averaging and the resulting  $\langle H \rangle_{est}$  is an average slope over all the ensembles for which the exponent is estimated and so is the standard error of estimation.

## 2.5 Analysis of DFA performance

The ability of DFA to yield  $H$  values intrinsic to a system was tested using the data of mBm. The most basic of which is to test how close  $\langle H \rangle_{est}$  is to  $\langle H \rangle_{act}$ . It is expected that  $\langle H \rangle_{est}$  would be sensitive to various parameters pertaining to the technique and of the data. Some of such basic parameters are:  $p$  the order of the polynomial,  $T$  the length of the data,  $R$  the range of  $H(t)$ , and, possible noise in  $H(t)$ .

We begin our numerical experiments by first testing to see if  $\langle H \rangle_{est} = \langle H \rangle_{act}$  whilst simultaneously observing the sensitivity of  $\langle H \rangle_{est}$  to  $p$  of DFA.

Time averaged Hurst exponent and its dependence on the order of detrending polynomial,  $p$

The first studies are designed to test how well a single exponent  $\langle H \rangle_{est}$  computed with DFA represents  $\langle H \rangle_{act}$  of the mBm data. For our numerical experiment, we chose mBm data with  $\langle H \rangle_{act}$  values ranging from 0.4 to 0.9 (steps of 0.05) with  $R = 0.06$ . The simulations are performed on mBm of size  $T = 2^{13} = 8192$  for one ensemble of the data with no ensemble averaging. This length of the data is chosen keeping in mind the nature of dependence of  $\langle H \rangle_{est}$  on  $T$  (discussed in the next section) and also to optimize the computational time.

Figure 2.3 shows the dependence of  $\langle H \rangle_{est}$  on  $\langle H \rangle_{act}$  for various detrending polynomial orders ( $p$ ) of DFA used. For  $p = 1, \dots, 5$ , DFA yields exponents close to

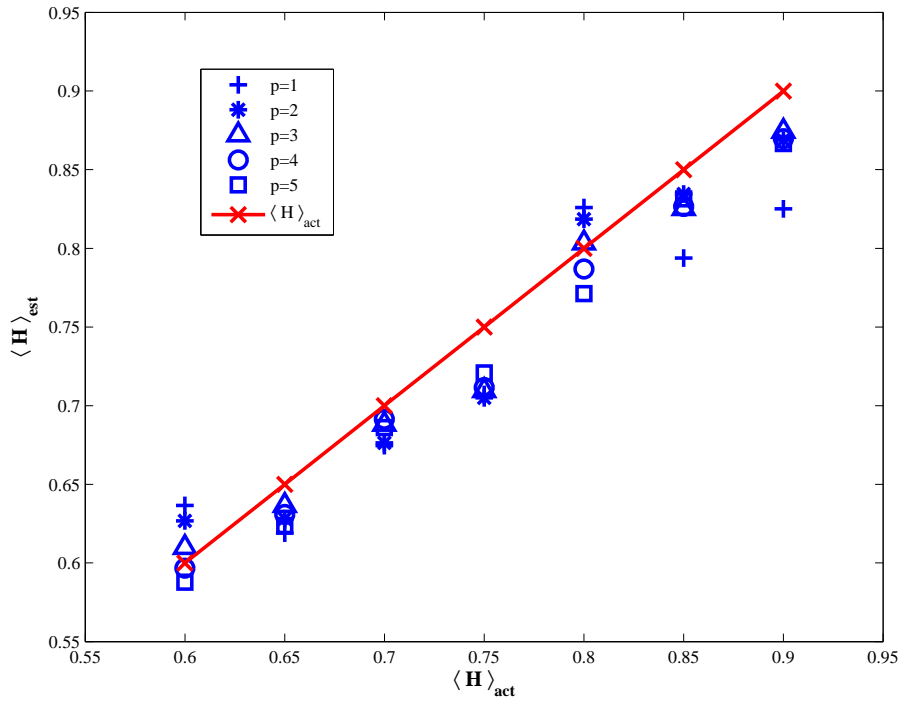


Figure 2.3: Dependence of  $\langle H \rangle_{est}$  on the polynomial order ( $p$ ) of DFAp used. The actual average values ( $\langle H \rangle_{act}$ ) are shown in red.  $R = 0.06$  for all  $H(t)$  used and  $T = 2^{13} = 8192$ . We see  $p = 1$  has higher  $EE$  compared to other cases.

$\langle H \rangle_{act}$ , albeit with varying accuracy. For each  $\langle H \rangle_{est}$ , from a DFA of order  $p$ , we see that there is no regular pattern of increasing/decreasing accuracy with increasing  $\langle H \rangle_{act}$ . This also concurs with our expectation that for a fixed  $R$ , the extent of LRC as measured by  $\langle H \rangle_{est}$  should not be dependent on  $\langle H \rangle_{act}$  of the data.

But for every  $\langle H \rangle_{act}$ , we see significant differences with in  $\langle H \rangle_{est}$  from various  $p$ . This is expected since lower polynomial orders ( $p = 1$ , linear) fit worse than higher orders because detrending using a linear fit on non-stationary data is expectedly worse, but this comes with a caveat that this is not necessarily so for higher order polynomials due to poor conditioning, particularly for lower time windows. This

may explain why we see that while  $p = 1$  results in higher error (lower accuracy) as compared to higher order  $p$ , there is no definite trend or difference in error for  $p = 2, 3, 4, 5$ . So, in the following studies, we use  $p = 2$ , i.e., a quadratic detrending which is also the most commonly used order.

To further settle on which length of data is appropriate within available computational resources, we proceed to test the effect of  $T$  on  $\langle H \rangle_{est}$  in the next section.

## Dependence on data size, $T$

Although this is not a strict condition, it is important to examine the data size effects on  $\langle H \rangle_{est}$ . This prompted us to set up the next experiment where we took various mBm data for which  $\langle H \rangle_{act} = 0.75$  and  $R = 0.06$  but of differing size  $T$ . Figure 2.4 shows this dependence of  $\langle H \rangle_{est}$  and the  $SE(\langle H \rangle_{est})$  on the size of mBm data ( $T$ ) used. Just like in the previous case, our result is for one ensemble of the data and so, there was no averaging. We fixed  $\langle H \rangle_{act}$ , to make our point in a more clear fashion as showing different  $\langle H \rangle_{act}$  would confuse one and may make it hard to see a clear pattern in  $\langle H \rangle_{est}$ .

We see that while longer data sizes do not necessarily effect the accuracy (difference between blue points and the reference line in red in Figure 2.4), it does improve the quality of estimation as seen by decreasing error bars - as measured by  $SE(\langle H \rangle_{est})$ . So, we fixed the length of the data to be  $T = 2^{13} = 8192$  for further simulations to assure lower  $SE$ .

The results so far help establish a benchmark on  $p$  and  $T$ , using which we can

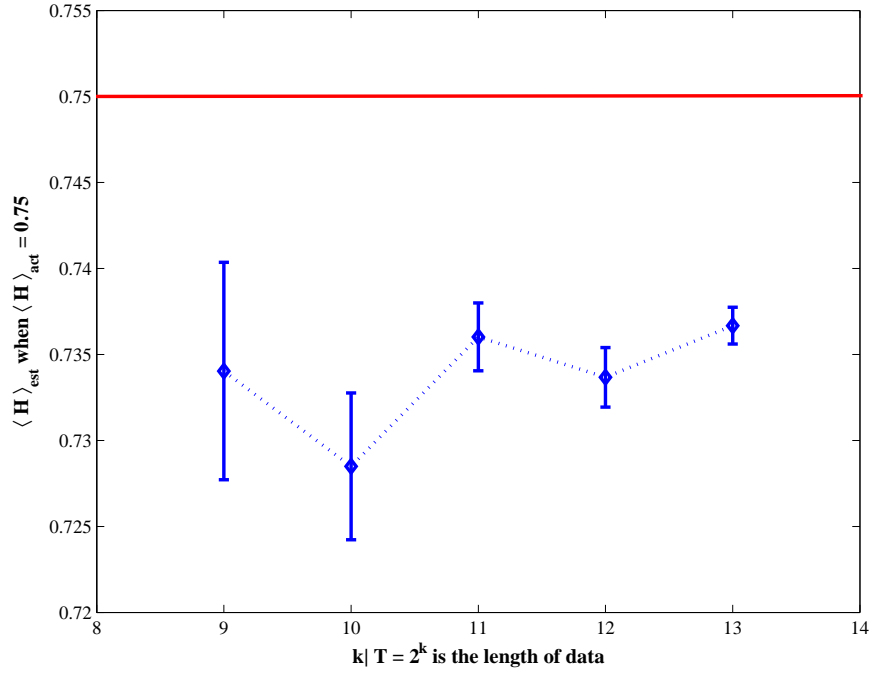


Figure 2.4: Dependence of  $\langle H \rangle_{est}$  on the length of data used for  $\langle H \rangle_{act} = 0.75$  and  $R = 0.06$ . It is interesting to note that while increasing data size has no predominant effect on  $EE$ , it does improve the quality of performance or decrease the  $SE$  as shown by error bars.

perform further tests. In the next section, we answer an important question: how  $R$  would effect  $\langle H \rangle_{est}$ ?

### Dependence on range $R$ of $H(t)$

This result is important because in the assumption of a local linear  $H(t)$ , one would like to know how the slope of the linearly varying  $H(t)$  as measured by  $R$  for fixed  $T$  and for each  $\langle H \rangle_{act}$ , would effect  $\langle H \rangle_{est}$ . To test this particular effect, we measure  $EE(\langle H \rangle_{est})$  for data with a  $\langle H \rangle_{act}$  (ranging from 0.4 to 0.9 in steps of 0.05) and  $R$  (values ranging from 0 to 0.1 in steps of 0.02). This accuracy is plotted



on a contour map as shown in Figure 2.5. The simulations were performed for 100 realizations of mBm data, each of size  $T = 8192$  and  $\langle H \rangle_{est}$  is estimated using DFA with detrending order  $p = 2$ . Hence, the resulting  $EE$  is an average over all the 100 ensembles.

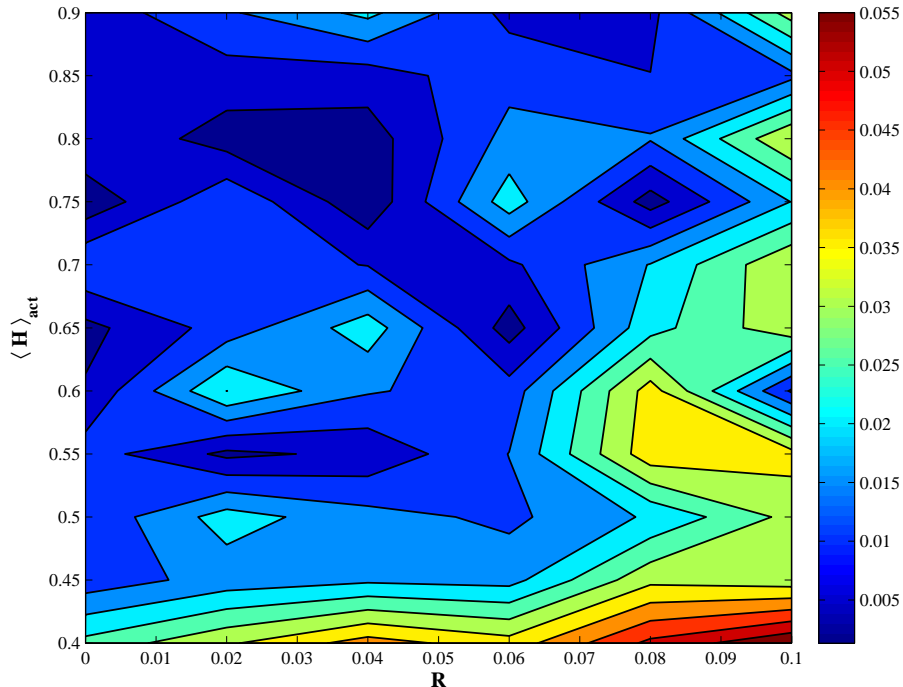


Figure 2.5: Contour map showing the dependence of relative accuracy of estimation (legend shown in color map on the right) on  $\langle H \rangle_{act}$  and  $R$ . We see interesting features emerging from this picture in that data with  $\langle H \rangle_{act} \geq 0.75$  have predominantly lower  $EE$  compared to other cases, and,  $R = 0.06$  serves as a transition point where  $EE$  increases rapidly beyond 2% approaching 5% near  $R = 0.1$ .

The ensemble averaging is used mainly because our results for a single sample of data were inconclusive. However, a clearer picture emerged as we averaged over several such ensembles. Our results show that there is a generally decreasing trend in accuracy (increasing  $EE$ ) with increasing  $R$ . This is in line with what we expect

when  $R$  (or the slope  $a$ ) increases with constant data size  $T$ , worsening performance (or increasing  $EE$ ) of DFA in estimating  $\langle H \rangle_{act}$  is expected intuitively. This argument is based upon the premise that when we enforce through the technique to estimate a single exponent from a data with a spread of exponents, we are actually averaging all the different possible Hurst exponents. With increasing  $R$ , the spread in various possible exponents, or the standard deviation increases and results in worsening accuracy as we are essentially trying to fit a broadly distributed (and possibly skewed) set of exponents to a narrow Gaussian distribution. This results in worsening accuracy, and, we stop the simulations at  $R = 0.1$  as this is the point where in general the accuracy starts to fall below 95% or  $EE \geq 5\%$  (as seen in ratio in the contour color map in Figure 2.5). We chose to use 5% as the limiting condition so as to standardize the benchmarks. Although it might sound arbitrary, this is a meaningful assumption because data that result in higher  $EE$  would typically show non-convergent behavior in their respective fluctuation plots for DFA, making it almost look like there are scaling crossovers which is a different phenomena altogether. Also, this result motivated us to use  $R = 0.06$  for our simulations in previous sections, since for this value of  $R$ , although the data is multifractional in principle, with accuracy in estimation over 98% for most  $\langle H \rangle_{act}$ , the data looks mono fractional in the view of the DFA technique.

There is another interesting result in Figure 2.5 which was not discussed in Figure 2.3: with increasing  $\langle H \rangle_{act}$ , over all  $R$ , the  $EE(\langle H \rangle_{est})$  is in general decreasing till  $\langle H \rangle_{act} = 0.75$  and then remains relatively invariant thereafter. This is interesting because it contradicts our earlier conclusion on the dependence of  $\langle H \rangle_{est}$  on  $\langle H \rangle_{act}$ .

It can be reasoned out that with increasing  $\langle H \rangle_{act}$ , as the level or correlations i.e., extent of LRC is increasing, DFA could do a better job estimating  $\langle H \rangle_{act}$  with a threshold occurring at  $\langle H \rangle_{act} = 0.75$ . However, this is an interesting phenomenon because data with  $H \geq 0.7$  is known to show predominant LRC behavior than the condition  $H \geq 0.5$ , which is commonly referred to as the Hurst effect. In data with  $\langle H \rangle_{act} \geq 0.75$ , and,  $R \leq 0.1$ , the  $H(t) \geq 0.7$ , which might imply a change in intrinsic nature of the data as speculated previously [Li and Zhao 2012] when compared with when  $0.5 \leq H(t) \leq 0.7$ . This motivates us to use  $\langle H \rangle_{act} = 0.75$  for cases in previous sections as this has minimum estimation error over all  $R$ , as seen in Figure 2.5.

In the following section, we intend to answer an interesting question- How robust is the DFA with respect to the presence of noise in linear  $H(t)$ ?

### Dependence on noise in $H(t)$

The robustness of DFA in the presence of noise is analyzed using a linear  $H(t)$  with additive noise. We simulate mBm data using  $H(t) = at + b + \epsilon_t$ , where  $\epsilon_t$  is drawn from Normal distribution ( $\mathcal{N}(0, \sigma^2)$ ), and,  $a$  and  $b$  are as defined in Equation 2.8. We control the strength of additive noise using parameter  $\kappa$  such that the standard deviation,  $\sigma = \langle H \rangle_{act} \times \kappa = 0.75 \times \kappa$  for  $\langle H \rangle_{act} = 0.75$  and  $R = 0.06$ . Other parameters,  $p = 2$  and  $T = 8192$  are the same as before and we average over 100 realizations of the data for each case.

From the results in Figure 2.6, we see that  $\langle H \rangle_{est}$  is very sensitive to  $\kappa$ , particularly to those values greater than  $10^{-3}$  beyond which the relative accuracy falls

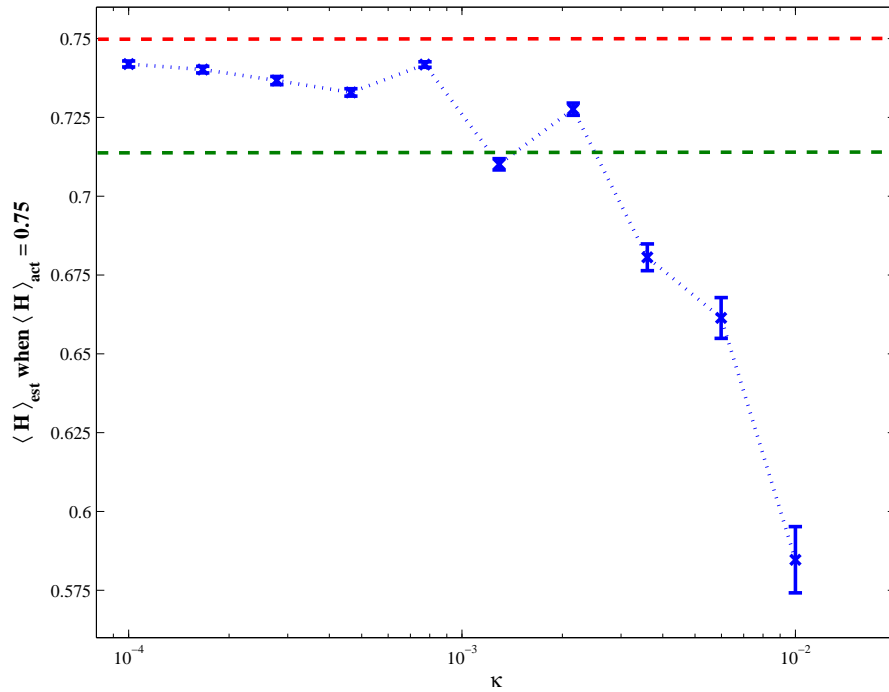


Figure 2.6: Dependence of  $\langle H \rangle_{act}$  on the strength of Gaussian noise ( $\mathcal{N}(0, \sigma^2)$ ) added to  $H(t)$ , represented here by  $\kappa = \sigma \times \langle H \rangle_{est}$ . We note how rapidly  $EE$  increases, well beyond the 5% limit (as shown in green line) as the strength increases beyond  $10^{-3}$ .

far below the 5% limit of  $EE$ . We can see the threshold with respect to the reference green line which represents 95% of  $\langle H \rangle_{act}$ . It is interesting to note that while it is expected to have worsening performance with increasing  $\kappa$ , the rate of loss of accuracy for  $\kappa \geq 10^{-3}$  is very high and goes to show how sensitive  $\langle H \rangle_{est}$  is to noise in  $H(t)$ .

## 2.6 Time varying Hurst exponents in AL index data

While mBm data models a more generalization over mono-fractal data, it is important to mention that AL index data from space weather also exhibits time

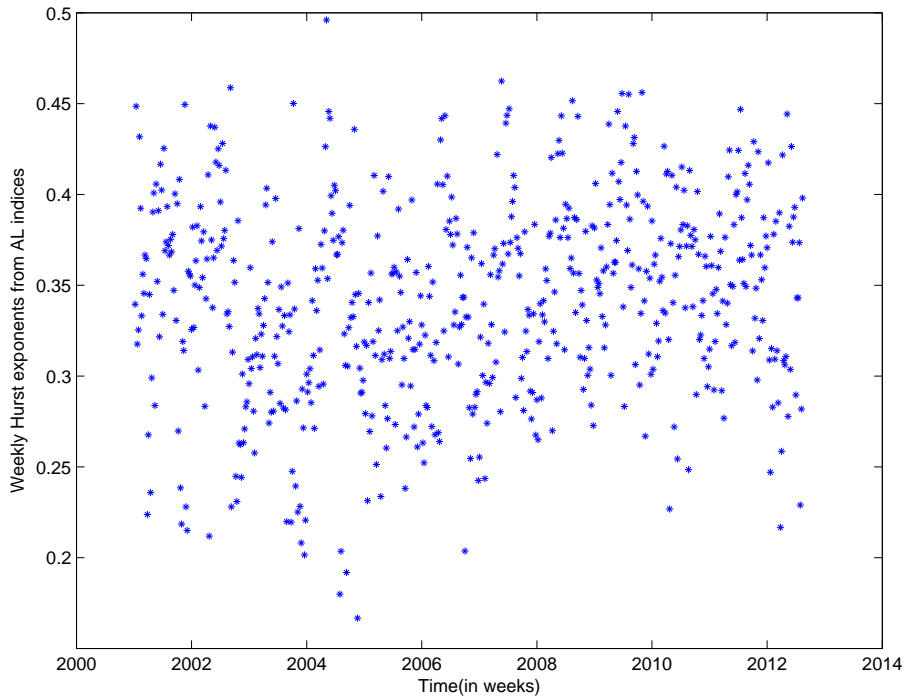


Figure 2.7: Time varying Hurst exponent picture of AL index data from 2001-2012 shown as weekly Hurst exponents.

varying fractal structure. This reflects in the time varying Hurst exponent picture of AL index. Our analysis estimates weekly Hurst exponents of AL index data from 2001-2012 and the results are seen in Figure 2.7. This makes a strong case on the relevance of this benchmarking given the multifractality of AL index data. Such multifractality was also found to exist in data from finance [Barunik et al. 2012, Bertrand et al. 2012]. In fact, models such as fBm, Levy, Truncated Levy were found to be incompatible with such multifractal features in financial data and hence, a multifractal alternative was proposed [Schmitt et al. 1999].

## 2.7 Conclusions

The characteristics of DFA in computing Hurst exponents from a non-stationary data with varying degrees of self-similarity (or LRC) are analyzed using mBm data. This is motivated mainly by increased use of DFA for real-life data, which more often than not is multifractional or with varying  $H$  [Hongre et al. 1999, Rogerio and Vasconcelos 2003, Wanlissm and Cersosimo 2006, Wei et al. 2004]. We set out with a basic and intuitively meaningful assumption that the estimated exponent within reasonably slowly varying  $H(t)$  is an average over time. This was shown to be true through numerical experiments on synthetic mBm data.

Furthermore, analysis of the effects of data and technique related parameters on the estimated exponent yielded some interesting insights and benchmarks for application of DFA on mBm. The following are the main conclusions on the usage of DFA technique for computing  $H(t)$  using sliding window estimation or just estimate a time averaged Hurst exponent:

1. The order of polynomial,  $p$  used for detrending in DFA has a limited and expected effect on our estimation as we see that even though mBm data is non-stationary, the non-stationarity is of a relatively simple kind and can be conveniently addressed by a lower order polynomial such as quadratic ( $p = 2$ ). Application of DFA with  $p > 2$  on mBm (and in general any data) should be done with caution due to poor conditioning of the higher order polynomial—more so in the lower time windows of estimation.

2. Data size (particularly  $T > 2^{11}$ ) does not predominantly affect the estimated exponent, although it would in general effect the precision (or standard error  $SE$  of least square fitting), more so due to limited number of averages performed in estimating the variance for fluctuation functions.
3. The range of the data,  $R$  which for a fixed data size controls the slope of  $H(t)$  has a major effect on the accuracy of estimation. We find that  $R = 0.06$  i.e.,  $H(t) = \langle H \rangle_{act} - \frac{R}{2} \rightarrow \langle H \rangle_{act} + \frac{R}{2}$  is a limiting condition for various data irrespective of  $\langle H \rangle_{act}$ . This conclusion is drawn on a 5% tolerance in  $EE$  as the limiting condition.
4. Increasing strength of additive noise diminishes the accuracy of estimation for obvious reasons. However, it is interesting to note how sensitive estimated exponents are with respect to the strength of noise. For a 5% tolerance in accuracy, we would desire that  $H(t)$  have no more than  $\langle H(t) \rangle \times 10^{-3}$  as standard deviation in noise. A noise of greater strength would quickly degrade the accuracy of the estimated exponents.

Beyond our observations on the nature of estimated exponents, we could also see a peculiar behavior of how DFA does a better job of estimating the time average of exponents when the data is predominantly self-similar ( $H(t) \geq 0.7$ ). This phenomenon although interesting would require further study.

With the knowledge gained through this study about DFA technique and its applicability, we now proceed to apply it to space weather data i.e., particularly the AL index data and study the following results in the next chapter.

## Chapter 3: Modeling crossover phenomenon in Detrended Fluctuation Analysis of Auroral Electrojet index data

### 3.1 Overview

In the last chapter, we introduced Detrended Fluctuation Analysis (DFA) technique and studied its robustness when applied to a general multifractal model - the multifractional Brownian motion (mBm). We also presented evidence of time varying fractal nature of Auroral Electrojet index (AL) data similar to mBm model (in Figure 2.7). Section 1.2 covers essential features of space weather, magnetospheric dynamics and AL index. Besides its time varying fractal nature as seen in Figure 2.7, AL index also has scaling crossovers in its fluctuation functions from DFA. Characterization of the crossover, modeling AL with the appropriate model that shows crossover features and using the aspects of model to estimate important parameters in AL data is covered in this chapter.

From previous work [Takalo and Timonen 1994], it was seen that the characteristic time scale of the AL index is about 30 minutes from auto-correlation estimation based study. Also, earlier work [Tsurutani et al. 1990] detected a crossover or a change in the spectral index of the Fourier components which limited the results (as



shown in Figure 3.1). The time scale corresponding to the frequency of crossover in power spectral density was estimated to be 5 hours. However, this crossover time was artificially selected and thus complicates subsequent analysis. There was no simple scheme to find the point of crossover when the data exhibits a single crossover in its spectral scaling or fluctuation analysis.

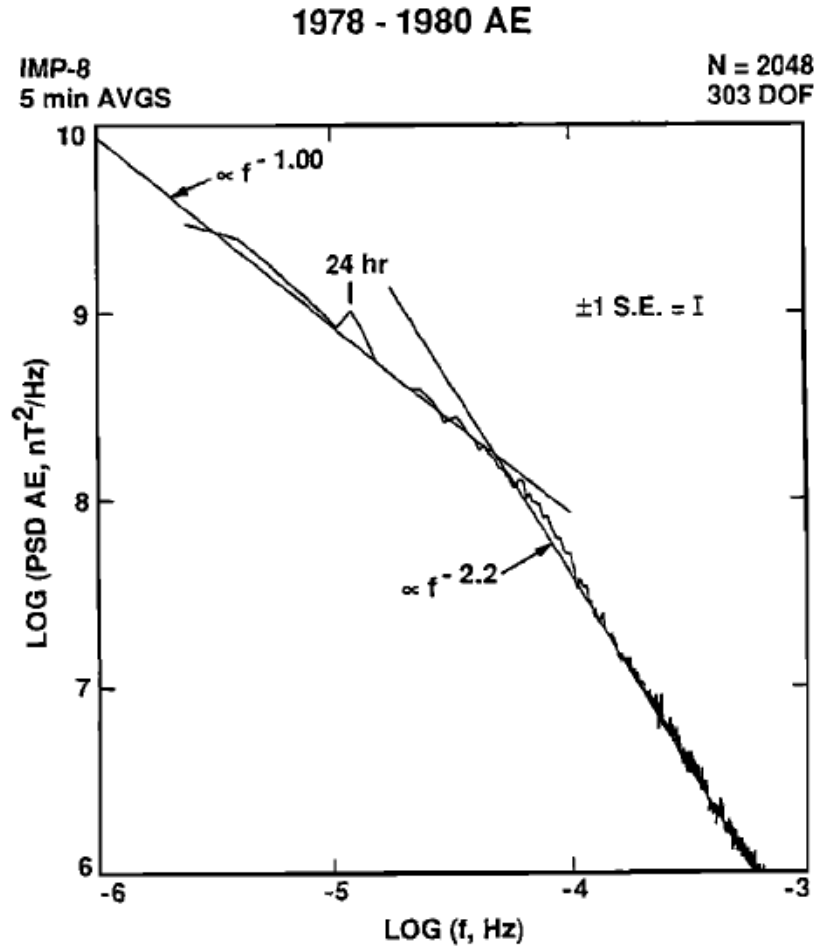


Figure 3.1: Power spectrum of 5 minute AE index from 1978-1980, we see a break/crossover in the spectrum at the frequency corresponding to a time of about 5 hours [Tsurutani et al. 1990].

Recently, fluctuation analysis using DFA was performed on AL index data [Sharma

and Veeramani 2011] where existence of crossover in fluctuation functions with two exponents was first observed. We replicated these results and extended the analysis using averaging over multiple ensembles of AL index data. Furthermore, we tested the dependence of exponents and the correlation crossover time scale on the order of polynomial used for detrending in DFA. In estimating the correlation crossover time scale and the Hurst exponents, we prescribe to an advanced regression to model two exponent power laws using Hyperbolic regression.

Recent work modeled AL index data using stochastic models [Anh et al. 2008, Rypdal and Rypdal 2010]. These models utilized the property of long range correlations seen in AL index. These models were successful in explaining the intermittency seen in AL data. However, these stochastic models do not explain the crossover seen in fluctuation analysis. We identified three theoretical models-two deterministic and one stochastic that all demonstrate such a crossover in DFA. Studying the nature of crossover seen in successive derivatives of these models and comparing the same in AL data, we could identify the stochastic model - Ornstein Uhlenbeck Langevin (OU-Langevin) best models the crossover effect. The correlation crossover time scale in OU-Langevin is known to depend on an important parameter in the model - the dissipation rate. This parametric significance explains the importance of robust estimation of the crossover time scale from AL data.

Estimation of the crossover time scale is an important step towards stochastic modeling of AL data using OU-Langevin model as mentioned earlier. However, estimation of correlation crossover time scale using DFA is riddled with challenges due to the additional detrending operation performed while estimating the variance [Havlin

et al. 1999]. The variance of OU-Langevin can be estimated analytically using multiple trajectory and single trajectory ensemble approaches which are described in Section 3.4. Using multiple trajectory ensemble averaging the correlation crossover time in AL data is estimated to be 4 hours 25 minutes which is close to the 5 hours crossover time in earlier work.

We begin by showing results from application of DFA on AL data in Section 3.2. Then, we present three theoretical models exhibiting crossover in DFA and present a basis for choosing OU-Langevin as closest model for AL data in Section 3.3. Next we demonstrate estimation of the dissipation parameter and hence, the correlation crossover time scale using variance estimation by modeling as OU-Langevin in Section 3.4 followed by conclusion in Section 3.5.

## 3.2 Fluctuation Analysis and scaling crossovers in AL index data

DFA estimates fluctuation functions for a given length of walk (or time window here) and under the assumption of scale invariance of these fluctuation functions estimates an exponent,  $\alpha$  as the scaling exponent. This scaling exponent  $\alpha$  and the Hurst exponent,  $H$  relate as:  $\alpha = H$  for stationary data, and,  $\alpha = H + 1$  for a non-stationary data whose incremental process is stationary. For example, a non-stationary fBm data is obtained with stationary fGn as its incremental process. This relationship would help relate  $H$  exponent to an exponent  $\alpha$  estimated from DFA technique in the next sections. These ideas have been previously discussed in Section 2.3. DFA is a step up from FA in that it estimates the variance in random

walk, detrended in the respective time windows (boxes) using a polynomial fit. The order of this polynomial ( $n$ ), used for detrending is fixed and the resulting technique is generally labeled as DFA $n$ . This  $n$  controls the extent of detrending performed in DFA technique and hence affects the estimated fluctuation functions.

For our analysis, the fluctuation functions from DFA $n$  (for  $n=1, 2, 3, 4, 5$  &  $6$ ) are estimated for time windows logarithmically equi-spaced between  $10$  to  $10^4$  minutes. The AL data was prepared in consecutive windows of  $10$  weeks ( $10^5$  data points in each window) of  $1$  minute AL data and there were total of  $100$  such ensembles in data from  $1980$ - $2013$ . Average fluctuation functions were estimated for the respective time window as an ensemble average for the  $100$  ensembles.

Figure 3.2 shows the results for DFA of AL for various orders of detrending polynomials  $n$ . We see that increasing  $n$  implies decreasing fluctuation functions as expected. Furthermore, the plots all converge with increasing  $n$ . This can be explained by the fact that higher order polynomial detrending is redundant after a certain minimum value of  $n$  taking into account the nature of non-stationarities seen in data.

We do see evidence of crossover in all the fluctuation functions. This implies we need a method to identify the point of crossover so that we can measure the exponents before and after the crossover. For a system that exhibits scaling crossover with one unique crossover point, there are many ways one can estimate the scaling before and after crossover and the crossover time. However, many of these techniques involve imposing some form or other of tolerance levels by the user. Hence, we suggest a method that limits such ambiguities and statistical artifacts. This

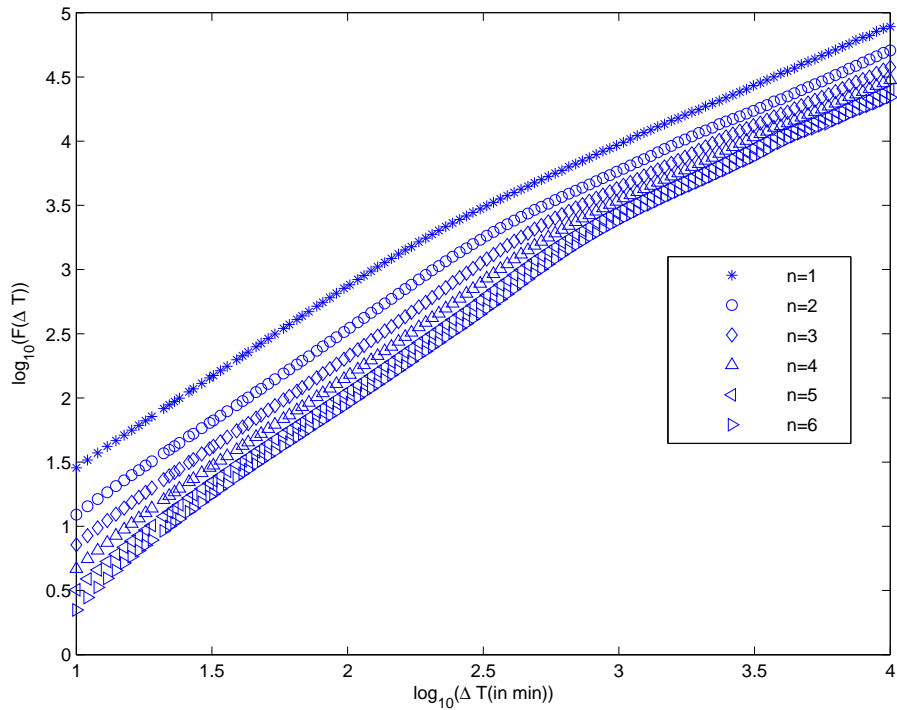


Figure 3.2: DFA of AL index- Fluctuation plots for AL index of 10 weeks in length averaged over 100 ensembles. DFA for various orders of detrending polynomial,  $n = 1, 2, 3, 4, 5$  &  $6$  is shown to illustrate the dependence of fluctuation plots on  $n$ .

method models the fluctuation function exhibiting crossover as a hyperbola with the respective arms of disparate scales approaching the two asymptotes of the hyperbola and the center of the hyperbola denoting the point of crossover.

A given data which exhibits crossover can be fit to a hyperbola following the method of least square [O’Leary and Zsombor-Murray 2004]. This method is based on quadratic constrained least-mean-square fitting and yields a best fit hyperbola for a set of scattered data. From the parameters corresponding to a standard hyperbola, one can obtain the slopes and intercepts of its asymptotes. Important advantage of this method is that it is numerically efficient to be applied to large data sets or large

ensembles of fluctuation functions in our case. Also, this method does not assume the point of crossover and only fits the data to the best possible hyperbola allowing us to determine if the nature of fit qualifies the data to be modeled as a hyperbola or not.

Furthermore, hyperbolic regression technique allows all of the data to model the nature of change in scales with asymptotic fits before and after the point of crossover. The slopes of the asymptotes are considered the Hurst exponents before and after the crossover respectively and the center of the hyperbola would be considered the crossover time from this technique. It is to be noted that the crossover point is not an exact time but a measure of where the change in scale takes place because more frequently the crossover or the transition in the scales takes place over a finite span of time where we can consider the two scales to be mixed.

Applying the hyperbolic regression to fluctuation functions of AL index data in Figure 3.2, we measure the respective Hurst exponents and crossover times. The estimated Hurst exponents for all  $n$  are as follows:  $\alpha_{shortterm} = 1.43 \pm 0.007$  or  $H_{shortterm} = 0.43 \pm 0.007$  and  $\alpha_{longterm} = H_{longterm} = 0.93 \pm 0.009$ . The resulting asymptotes and their slopes are shown in Figure 3.3. These estimated exponents imply that the nature of correlations in the data change from non-stationary uncorrelated (anti-correlated) to stationary highly correlated through the crossover regime.

It can be also be seen from Figure 3.3 that the crossover time changes between DFA1 and DFA2 from 224 minutes (3 hours 44 minutes) to 420 minutes (7 hours). This has been seen to be true with other  $n$ , i.e., the crossover time  $T_{crossover}$  increases

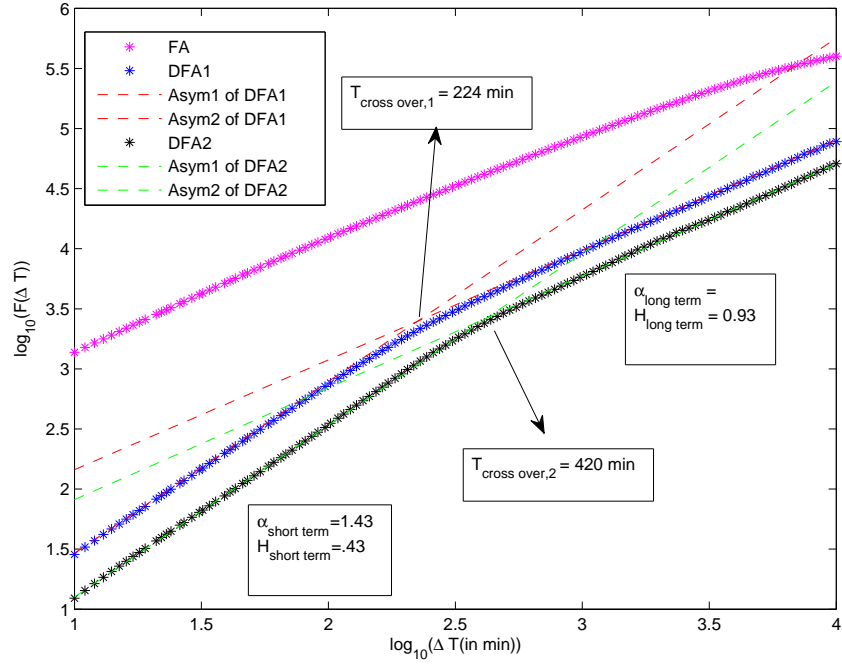


Figure 3.3: Hurst exponent estimation from FA, DFA and cross over point detection using Hyperbolic least square fitting for AL data. FA fails to estimate the correct Hurst exponent before the crossover due to the non-stationary nature of AL data. DFA1 and DFA2 do a better job but show different crossover points.

with increasing  $n$ . From plots in Figure 3.2, we see that the fluctuation functions at lower values of window lengths i.e.,  $\Delta T \leq 100$  minutes before the crossover time do not converge as much as those at higher values of window lengths i.e.,  $\Delta T \geq 1000$  for all  $n$ . The two asymptotes from DFA1 and DFA2 from Figure 3.3 are parallel to each other and this is true irrespective of  $n$ . However the crossover time increases with increasing  $n$ . Due to this, the fluctuation functions converge more at higher window lengths than those at lower window lengths as observed from Figure 3.2.

Hence, it is now established that estimation of long term and short term Hurst

exponent is not effected by  $n$ , but the crossover time changes with  $n$ . Due to this, we select DFA1, the DFA which least effects the data in terms of detrending operation (removing linear trends) for further analysis. Solar wind (given by  $-v.B_z$  using bulk velocity  $v$  of solar wind and magnetic field  $B_z$  of IMF) is known to be the driver for magnetospheric dynamics as described in Section 1.2. However, magnetospheric response is known to demonstrate internal dynamics with a time scales of its own [Baker et al. 1997]. In order to correlate and understand the crossovers better, we compare the fluctuation analysis from DFA1 of AL and solar wind data as seen in Figure 3.4. The solar wind data is obtained over a span of 12 years from September 2000 to February 2013 which contains roughly  $10^6$  time steps at 5 minute intervals.

We see from Figure 3.4 that there is a scaling crossover in solar wind data too and it occurs at about 1 day as estimated from the hyperbolic regression. The time scale at which crossover in solar wind data is seen i.e., 1 day is much larger than the time scale of crossover of AL data i.e., 5 hours. This implies that solar wind is primarily single exponent in nature in comparison to time scales relevant to AL data. However, the Hurst exponent from solar wind before its crossover time scale at 1 day ( $\approx 0.9$ ) is estimated to be close to the Hurst exponent from AL data after its crossover time scale at 3 hours 44 minutes ( $\approx 4$  hours) from DFA1 ( $= 0.93$ ) indicative of the driving nature of the solar wind.

DFA technique combined with hyperbolic regression has so far proven to be an effective tool to estimate Hurst exponents in a two exponent multi-scale system such as data from AL index. However, we have seen how  $n$  the order of detrending



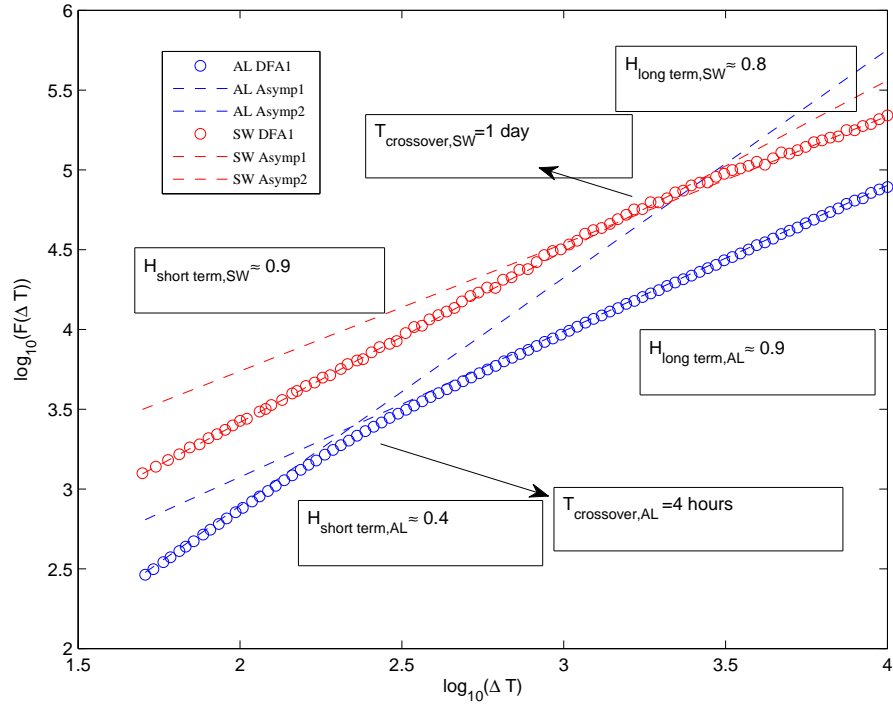


Figure 3.4: DFA1 of AL (in blue) and solar wind data (in red). We see that the long term Hurst exponent of AL is equal to the short term Hurst exponent of solar wind data.

polynomial severely affects the correlation crossover time limiting our ability to robustly estimating it. Several theoretical models demonstrate similar crossover in fluctuation functions from DFA. In the next section, we present evidence of three such models and propose a test to benchmark their crossover features to help us chose the best model for AL data.

### 3.3 Theoretical models for crossover phenomena

Time series data from various theoretical models exhibit crossover in fluctuation functions and can be used to model AL data. Most basic are periodic systems

like sine:  $x(t) = A\sin(2\pi\omega t)$ , where  $A$  is amplitude and  $\omega$  its frequency, and strange attractors like the Lorenz attractor (which was described in Section 1.5). It is to be noted that  $1/\omega$  is a measure of crossover time scale as we show in results from further analysis. Both of these models are deterministic in nature.

Stochastic systems like OU-Langevin process [Chandrasekhar 1943, Langevin 1908, Uhlenbeck and Ornstein 1930, Wang and Uhlenbeck 1945] also demonstrate crossover in their fluctuation analysis. The stochastic differential equation describing OU-Langevin process is:

$$\frac{dx(t)}{dt} = -\lambda x(t) + \eta(t) \quad (3.1)$$

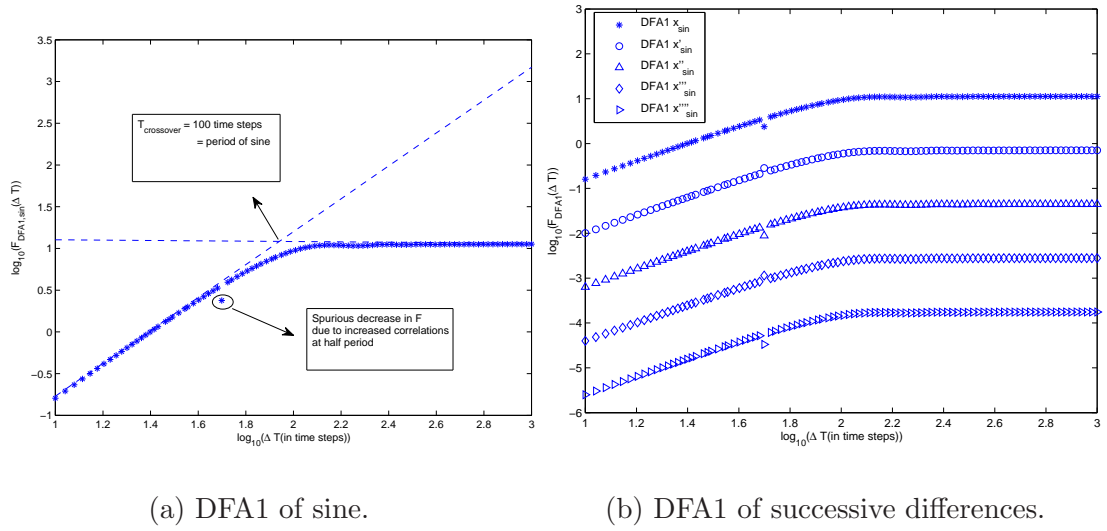
where  $\eta(t)$ , the stochastic driver, is Gaussian noise with 0 mean,  $\sigma$  standard deviation and  $\lambda$ , the dissipation parameter or the coefficient of friction in the drift term. It is to be noted that  $1/\lambda$  is measure of crossover time as we show in our results soon. The Ornstein Uhlenbeck process is primarily used to model the velocity of a massive Brownian particle suspended in a fluid. This model overcomes, at least the difficulty encountered if we try to model the location of the particle by Brownian motion, which results in infinite velocities.

In order to identify the best model that explains the nature of crossover seen in AL data, we need to test using DFA1 the nature of crossover in these three models. However, all the three models show decreasing LRC nature i.e., decreasing Hurst exponents through the crossover. Hence, DFA1 of the time series data corresponding to these models alone cannot help us conclusively to identify suitable model. So, we propose to test the DFA1 of successive derivatives or the differences (in discrete

time) to identify in depth the signature features corresponding to scaling crossover seen in these models.

Figure 3.5 shows the result for DFA1 on sine data. In particular, Figure 3.5a shows the result from application of DFA1 on time series corresponding to sine with a frequency of  $1/100$  or  $1/\omega = 100$  time steps. As we see, DFA1 is successful in identifying the crossover time for sine data. The amplitude of the sine wave does not hold relevance in our analysis because it does not control the crossover time. Hence, we arbitrarily chose a value of  $A = 10$  for the time series data for sine. In Figure 3.5b, we see the results for DFA1 of successive differences in sine data. Derivative of sine is cosine and vice versa. DFA1 of cosine is of similar nature to DFA1 of sine as the nature of LRC changes at the time scale of  $1/\omega$  which remains the same in the process of taking the derivative or difference. Hence, we see that with the exception of a constant decrease in fluctuation functions from DFA1, the nature of crossover exhibited remains the same in time series successive differences of sine data.

Figure 3.6 shows the result for DFA1 on x-component of the Lorenz data. Figure 3.6a shows the result from application of DFA1 on time series corresponding to Lorenz. The system is integrated (using system of differential equations with standard parameters as mentioned in Section 1.5) such that the correlation time scale as measured from mutual information is roughly 15 time steps of integration. This is an estimate of the time scale over which correlations are completely lost within the data. We see that DFA1 is not successful in identifying the crossover time for Lorenz data due to the detrending operation. However, we interpret the



(a) DFA1 of sine.

(b) DFA1 of successive differences.

Figure 3.5: DFA1 of sine and its successive differences.

crossover time as the time required to traverse a wing of the Lorenz attractor before it switches to a different wing.

The asymptotic nature with two exponents is not seen within the time span over which DFA1 is estimated. In Figure 3.6b, we see the results for DFA1 of successive differences in Lorenz data. Due to nonlinear nature of the system, the crossover is seen in DFA1 of successive differences. We noted that the fluctuation functions reach a noise floor at the 6th difference but this is not shown in the Figure as it is not relevant for our analysis. We consider these features seen particularly from result in Figure 3.6a as the signature features of the crossover in data from Lorenz attractor.

Figure 3.7 shows the result for DFA1 on integrated data from the differential equation corresponding to OU-Langevin mentioned in Equation 3.1. Figure 3.7a shows the result from application of DFA1 on time series corresponding to solution of

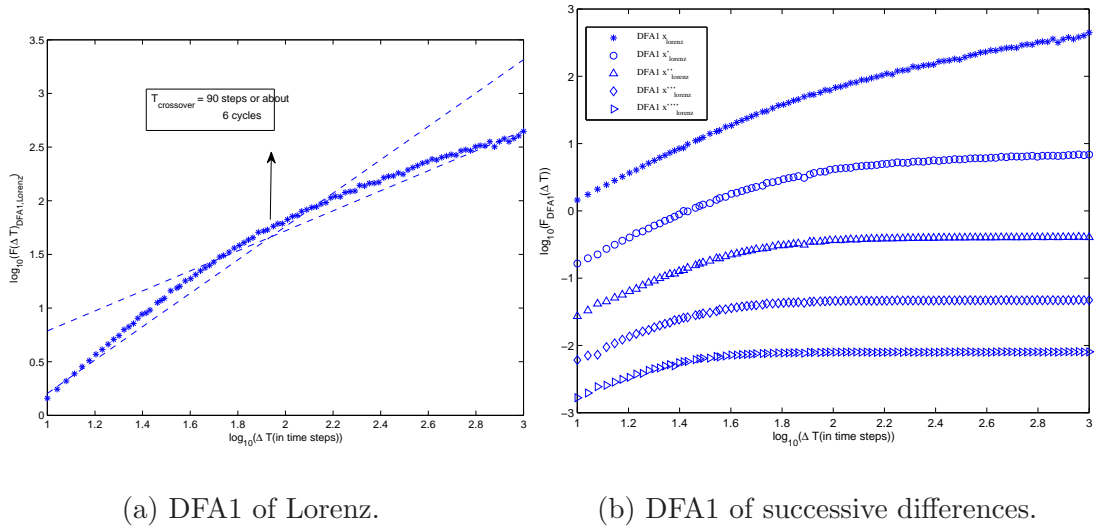
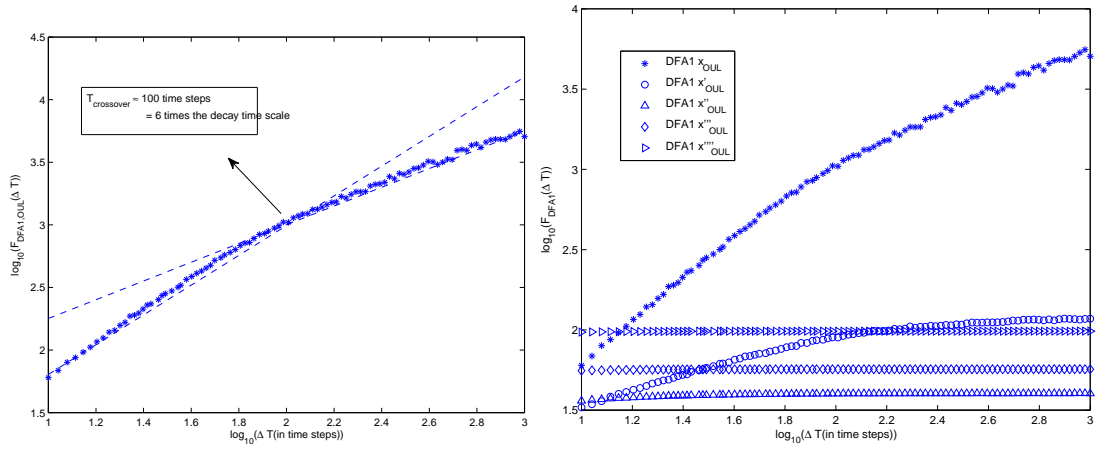


Figure 3.6: DFA1 of Lorenz and its successive differences.

OU-Langevin. The system is integrated using an initial condition  $x(0) = 0$  and with a dissipation parameter  $\lambda = 0.055$  such that crossover time is  $1/\lambda = 18$  time steps of integration generating discrete time series. The noise added to the differential equation is drawn from gaussian distribution or normal distribution with 0 mean and a standard deviation of  $\sigma = 40$ . Due to detrending using linear fits done by DFA1 method, the crossover time is not estimated exactly as seen in Figure 3.7a.

Figure 3.7b shows the result for DFA1 of successive differences of solution of OU-Langevin equation. We see that the first difference (or derivative in continuous time) exhibits crossover in fluctuation functions. Further orders of difference do not show crossover in fluctuation functions and approach a noise floor for the second difference. In fact, this reflects in the increasing fluctuation functions from third difference onwards. This is owing to the fact that the incremental process of gaussian noise would have a higher standard deviation than the gaussian noise reflecting in



(a) DFA1 of OU Langevin.

(b) DFA1 of successive differences.

Figure 3.7: DFA1 of OU-Langevin and its successive differences.

increased variance as estimated through fluctuation functions.

Figure 3.8 shows DFA1 on successive differences in AL data from which we see strong similarity in the nature of crossovers in fluctuation functions from second difference as seen previously from Figure 3.7b for solution of OU-Langevin. The results from DFA1 of successive differences, show that OU-Langevin best explains the feature seen in AL data. Hence, using this as basis, we model AL data as OU-Langevin.

There is a driver term in the stochastic differential equation describing OU-Langevin process which is the additive gaussian noise  $\eta$  in Equation 3.1. The solar-wind input driving the magnetosphere acts similarly, which due to its turbulent nature is multifractal in nature. The frictional drift with dissipative term  $\lambda$  as coefficient of friction in Equation 3.1 represents negative feedback to the system. This feedback drift term can be seen trying to establish equilibrium or maintain

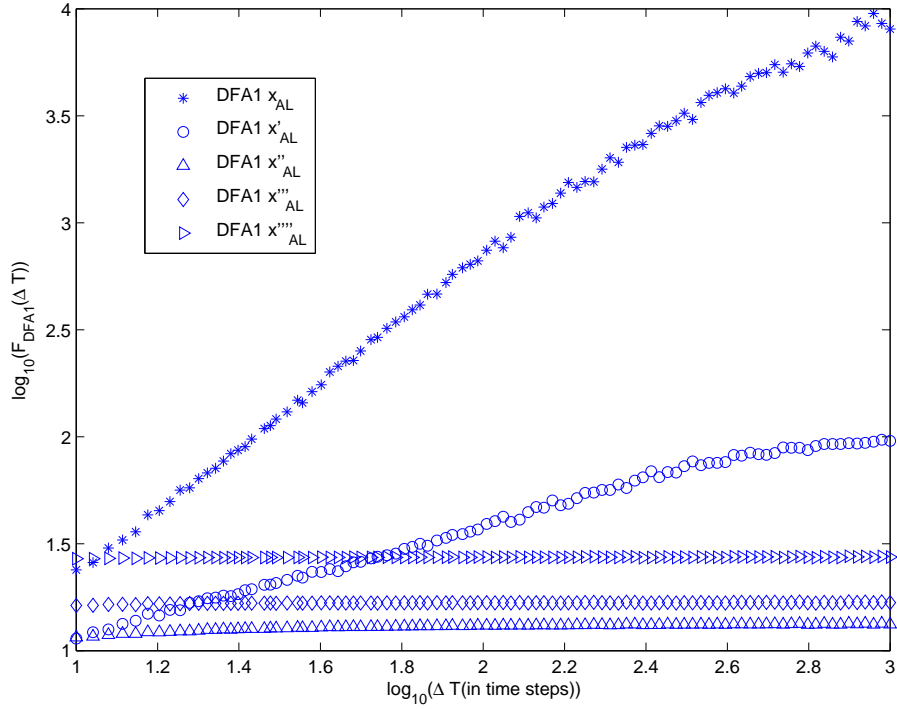


Figure 3.8: DFA1 of successive differences of AL index data.

stationarity in response to the additive noise. Hence, these similarities show that the magnetospheric dynamics can be modeled using OU-Langevin process.

Aided by battery of tests using DFA1, we could establish reasons for using OU-Langevin as the appropriate stochastic model for AL index data. The crossover time in OU-Langevin is known to be equal to the inverse of dissipative term. Modeling of AL data using OU-Langevin model helps us estimate the dissipative term from which we can estimate its crossover time. This crossover time, as discussed in the introduction is a measure of the time scale in which the correlation nature of the data changes. Hence, it is a measure of the time scale of the internal dynamics of the magnetosphere. In the next Section, we demonstrate application of OU-Langevin

model in estimation of the correlation crossover time scale of AL index data.

### 3.4 Stochastic modeling of AL index and estimation of crossover time

Fluctuation functions estimated by a fluctuation analysis method utilize a measure of spread in the random walks constructed using the data as the incremental process. DFA is necessary to estimate the fluctuation functions in data with non-stationarities. As AL data is non-stationary in nature at short time scales i.e., before the crossover time scale, we needed DFA to estimate the fluctuation functions. As seen in Section 3.2, DFA fails to estimate the correlation crossover time scale due to the detrending operation. Hence, the main intention behind stochastic modeling of AL data is to find an alternative method to estimate this crossover time scale.

Variance or spread in the data within a time window can also be measured without the random walk approach taken by fluctuation analysis techniques such as DFA. A variance estimated from one realization of the data is not a representation of the variance of the entire data. Hence, we need an ensemble average to estimate the mean variance. Of such ensemble averaging methods, Multiple Trajectory Ensemble (MTE) and Single Trajectory Ensemble (STE) averaging are most commonly seen. In MTE approach, averaging is performed over independent realizations of the data all starting with the same initial value i.e.,  $x(0) = 0$ . In STE approach, averaging is performed over realizations each of which starts are different initial points. This is because a single trajectory or data is divided into multiple windows each with differing initial points. Earlier, equivalence between MTE and STE approaches was



assumed without proof [Wang and Uhlenbeck 1945].

Recently, it was shown that variance estimated using MTE approach can be derived analytically for OU-Langevin process as [Ignaccolo et al. 2010]:

$$\sigma_M(t)^2 \equiv \langle x(t)^2 \rangle_M - \langle x(t) \rangle_M^2 = \frac{\sigma_\eta^2}{2\lambda} [1 - e^{-2\lambda t}] \quad (3.2)$$

where the subscript M refers to the MTE averaging method and  $x(0) = 0$  for all the ensembles.

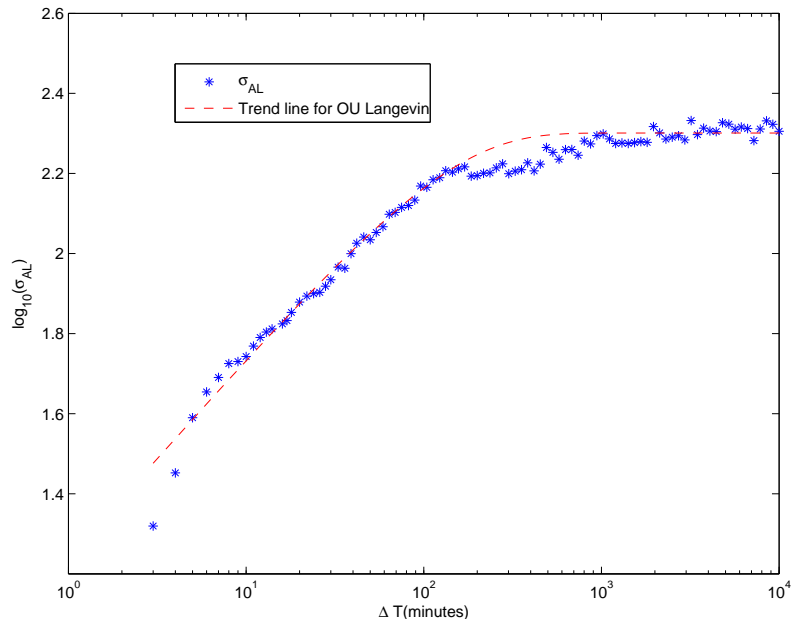


Figure 3.9: Variance of AL estimated from multiple trajectory ensemble (MTE) approach (in blue) and the theoretical curve from an OU-Langevin process with similar crossover time of 265 minutes or 4 hours 25 minutes (in red).

Estimating variance using MTE approach in time series data corresponding to physical processes is not always possible due to strict requirement of zero initial conditions i.e.,  $x(0) = 0$ . However, we could identify 100 independent windows all with initial conditions of AL values close to 0 nT and whose data length can be as

high as 1 week of 1 minute values ( $\approx 10^4$  points). This afforded us the flexibility of using MTE approach to estimate the variance from such time windows of AL data. We estimated the variance of AL data at several window lengths similar to fluctuation analysis by MTE approach. Results for this variance estimation are seen in Figure 3.9.

Modeling AL data using OU-Langevin process implies that we can estimate the dissipation parameter  $\lambda$  corresponding to the stochastic model. Equation 3.2 shows that for OU-Langevin model, the variance, as estimation using MTE approach is analytically exact. Hence, we fit the variance of AL index data to theoretically expected variance from OU-Langevin model obtained using MTE approach. This curve is shown in red dashed line in Figure 3.9. The crossover time from variance of AL index data is obtained using hyperbolic regression as 265 minutes (4 hours 25 minutes). Using this value of crossover time, we can calculate the dissipation parameter as  $\lambda = 1/265$ . Furthermore, fitting the variance to the theoretically expected variance from Equation 3.2 with  $\lambda = 1/265$ , we estimated the standard deviation of the gaussian noise as  $\sigma = 13.6$  nT.

### 3.5 Conclusions

We used fluctuation analysis by DFA technique to show the existence of two Hurst exponents in AL index data. Furthermore, we introduced an innovative and alternative regression scheme to model such crossovers in power law fits - hyperbolic regression. This regression scheme helped us estimate the short term and long term

Hurst exponents in AL data. However, due to shift in fluctuation functions due to detrending done in DFA technique, we cannot measure the correlation crossover time scale.

Various theoretical models show similar crossover in fluctuation functions from DFA. One such model is the stochastic model - the OU-Langevin process. Tests using DFA1 of successive differences in the time series data helped us establish the validity of OU-Langevin as the appropriate model to explain crossover seen in fluctuation functions seen in AL data.

Variance estimation using multiple trajectory ensemble (MTE) approach to estimate crossover time after modeling AL data as OU-Langevin helps us estimate accurately the crossover time. Results from such an analysis are shown and this is the first known such study which robustly estimates the correlation crossover time - a significant multi-scale feature in AL data. This crossover time is a measure of transition time scale between internal dynamics of AL data and that of driver solar-wind.

In this chapter, we modeled crossover feature in fluctuation analysis of AL data. For AL index data, there is no variation in the nature of crossover i.e., the LRC nature always decreases with time. However, this is not the same for other data sets from finance and physiology. In the following chapters, we will show that there are differing natures within crossover phenomenon across global financial index data or between different heart conditions respectively. We use our approach to model crossover phenomenon i.e., hyperbolic regression scheme to study these features and obtain a better understanding of magnetospheric dynamics.

## Chapter 4: Varying nature of crossover phenomenon across global financial markets

### 4.1 Overview

In the previous chapter we studied the crossover phenomenon seen in magnetosphere data i.e., AL index data. Recently there has been work identifying similar crossover features in other data i.e., data from finance and physiology. In this chapter, we take data of a set of global composite stock indices and perform crossover analysis by hyperbolic regression on them. Previous work [Carles 2000, Liu et al. 1999] shows that financial data exhibits scaling crossover phenomenon. In fact this feature has been used for better modeling of volatility within index data. Analysis over a more general volatility with a power of the absolute return also demonstrated crossover in fluctuation analysis [Shi-Hao 2009].

This motivated to use our regression scheme to characterize the crossover seen in their results. Also, we were motivated to apply the same on data from other markets different from the Asian markets used in their analysis. Hence, we used time series data corresponding to daily closing index values of DJIA (American Dow Jones Industrial Average), NASDAQ (American National Association of Securities Dealers

Automated Quotations), SSE (Chinese Shanghai Stock Exchange), N225 (Japanese Nikkei 225 index) and FTSE (UK based Financial Times Stock Exchange). We considered 2 US stock indices, 1 from China, Japan and UK each, respectively. These are used considering the fact that they have been around long enough to provide us with long enough data for fluctuation analysis and their distribution across various geographies on a global scale.

We chose these particular markets given the apparent differences in the nature of the financial markets in these countries. The level of control exerted on the local market structure and growth in each case is different. Each index represents markets that have two kinds of mechanisms driving it as discussed previously in Section 1.2. One macroeconomic i.e., business cycles within each company included in the index. The other is more microeconomic or the local investor speculation that is controlled by the local aspects such as performance of the company and sociopolitical influences. These two mechanisms may affect the low frequency and high frequency components of its power spectrum and thus is reflected in the crossover seen in fluctuation functions.

Our results throw light on some interesting features within such data. First of all, there is stark contrast in the nature of evolution of long range correlation (LRC) nature in the data across the crossover regime for different markets. Particularly, we see that the DJIA, NASDAQ and N225 indices demonstrate increased correlations with time, and, SSE demonstrates decreasing correlations with time. FTSE does not show predominant crossover phenomenon. This is one of the first such study on such differing natures of financial indices and can help understand the differing

nature of complexities in diverse markets, while finding regions of similar scaling between each other. Furthermore, we consider two theoretical models that capture such increasing (concave) and decreasing (convex) LRC natures and speculate on the nature of the markets using such models.

## 4.2 Scaling crossover in financial indices

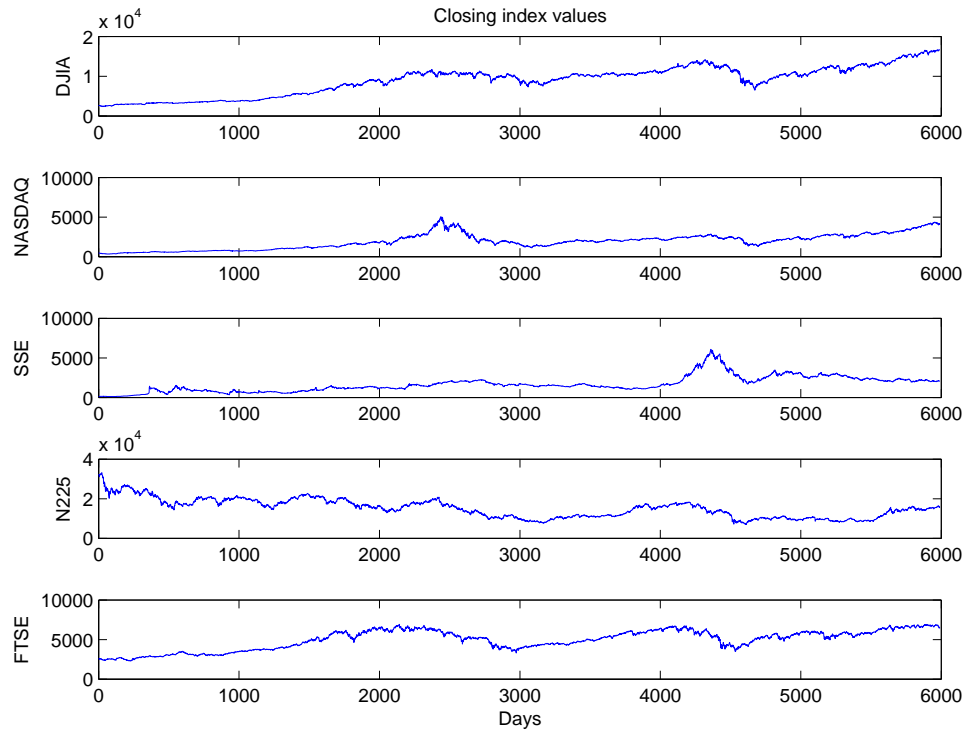


Figure 4.1: Daily closing prices for DJIA, NASDAQ, SSE, N225 and FTSE for 5990 days of trading days till April 27<sup>th</sup>, 2014.

The data from stock indices is as shown in Figure 4.1 with daily closing prices for DJIA, NASDAQ, SSE, N225 and FTSE for 5990 days all ending on April 27<sup>th</sup>, 2014. This length of the data is considered taking into account the length of shortest

data set-of that of SSE index. Given the closing stock price at time  $t$  (in days) as  $P(t)$ , the return of the stock index price,  $R(t)$ , after an interval  $\Delta t$  ( $=1$  day) is given as:

$$R(t) = \ln \frac{P(t + \Delta t)}{P(t)} \quad (4.1)$$

$|R(t)|$  is referred to as volatility and generalized volatility is given as  $|R(t)|^\gamma$ , where  $\gamma$  the power exponent that controls the weights assigned to small volatility (with  $\gamma < 1$ ) or large volatility (with  $\gamma > 1$ ) respectively.

Previous work proposed usage of the generalized volatility to control various crossover features such as the extent of crossover and the crossover time. We first show results from prior work setting context for our current analysis.

#### 4.2.1 Crossover in financial data and previous results

Previous work [Shi-Hao 2009] demonstrates using the generalized volatility data to illustrate crossover phenomenon. The results are shown in Figure 4.2. In Figure 4.2a, we see that DFA is applied to generalized volatility with various  $\gamma$  values. We see that the spread of scales or the difference in Hurst exponents across the crossover decreases with increasing  $\gamma$ . The power exponent controls the strength of small returns and amplifies them. The extent of crossover, as can be measured by DFA exponent  $\alpha$  is shown in Figures 4.2b decreases with increasing  $\gamma$ .

Fixing a particular choice of  $\gamma$  is important to illustrate difference in such crossover features across different markets. In our study, for all the indices used for this study, we fix  $\gamma = 0.1$  in estimating generalized volatility to maintain uniformity.

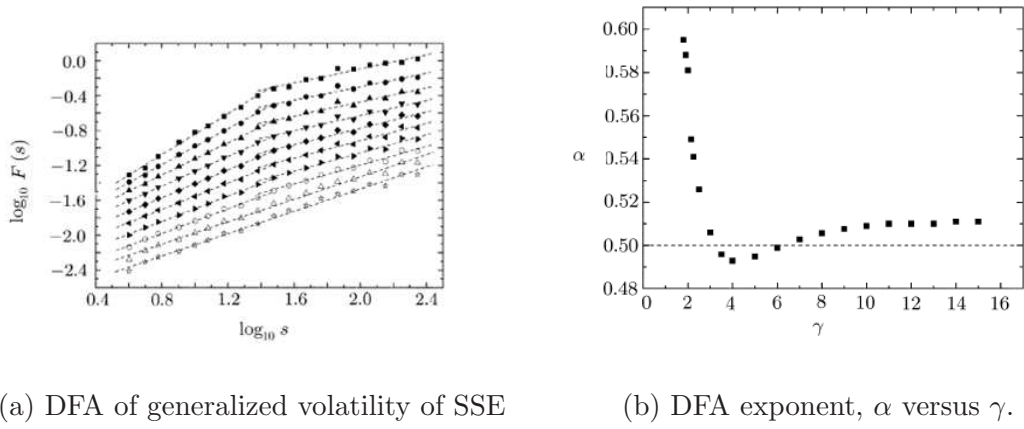


Figure 4.2: DFA of generalized volatility of SSE at various power exponents [Shi-Hao 2009] in order of increasing power exponents (ordered top to bottom) shown in (a). Also, the DFA exponent shows the separation of scaling across crossover and is shown in relation with power exponent in (b).

DFA is applied to generalized volatility of all the indices shown in Figure 4.1. Results of the analysis are discussed in the next section.

#### 4.2.2 DFA of financial indices and crossover analysis

We first perform DFAn with various polynomial orders  $n$  on generalized volatility ( $\gamma = 0.1$ ) of DJIA index to study two important factors - (1) would the crossover remain at around the same point if the length of the data is changed? and (2) what is the appropriate  $n$  for DFA to be applied?

Results of these analysis are seen in Figures 4.3 and 4.4. In Figure 4.3, DFAn with  $n=0, 1, 2, 3$  &  $4$  along with FA of generalized volatility in DJIA data from 1986 till 2014 are shown. Figure 4.4 contains the same analysis as applied to data from 1990-2014 i.e., the data corresponding to data size of the shortest of all indices



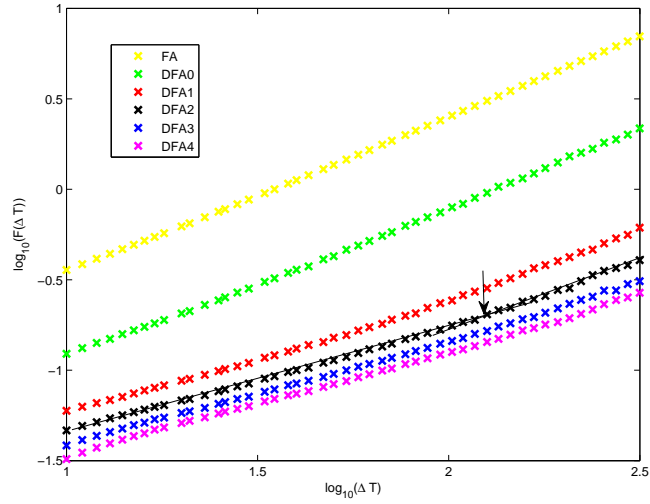


Figure 4.3: DFA $n$  and FA of DJIA generalized volatility for its complete data length (1986-2014). Note: window size  $\Delta T$  is in days.

used-SSE. We see that different orders of detrending polynomial affect the results.

Primarily, we note that FA and DFA0 do not show the crossover in fluctuation functions due to their inability in handling non-stationarities. We begin to see the crossover phenomenon for  $n \geq 1$ . Also, plots from DFA3 and DFA4 are close, thus indicating convergence. Usage of higher polynomial particular  $n \geq 3$  complicates the estimation of fluctuation functions because of artifacts introduced by over detrending within the windows. This results in additional crossovers in fluctuation functions which are not due to the data used. Hence we use DFA2, considering all the effects, for further analysis on other indices.

Also, we see that the crossover time as shown by the black arrows in Figures 4.3 and 4.4 are 100 and 120 days, respectively. This crossover time is estimated using hyperbolic regression scheme described in the previous Chapter. It is interesting to note that although the data used for Figure 4.4 is a subset of data used for Figure 4.3,

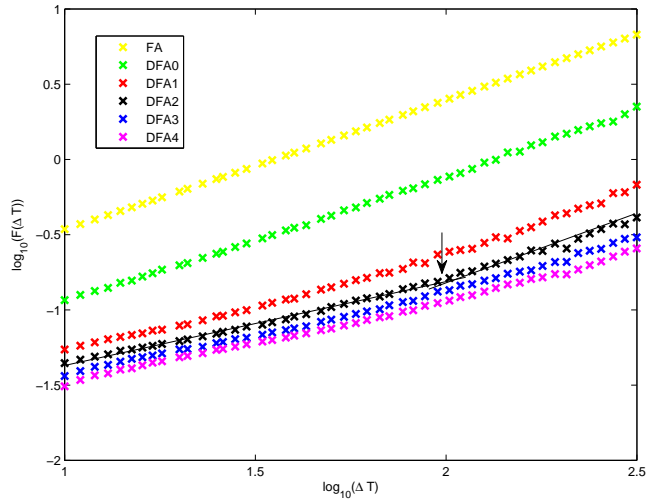


Figure 4.4: DFAn and FA of DJIA generalized volatility from 1990-2014. Note: window size  $\Delta T$  is in days.

the crossover time is nearly unchanged. This suggests that the features in subsets of data are maintained as in the whole data. This is a desirable result since we are truncating the data for other indices to compare data from equal span of time. It should be noted that we intend to compare the nature of crossover in fluctuation functions in each stock index. Due to dependence of the Hurst exponents on the power exponent,  $\gamma$ , and the crossover time on the order of detrending polynomial in DFA, we do not estimate these parameters.

The features mentioned earlier in our analysis on DJIA are seen in DFAn of NASDAQ and SSE also, as shown in Figures 4.5 and 4.6 respectively. However, the primary difference between NASDAQ and SSE is that the extent of LRC decreases in SSE where as increases in NASDAQ data after the correlation crossover time. This is an interesting difference which we discuss further in the next section.

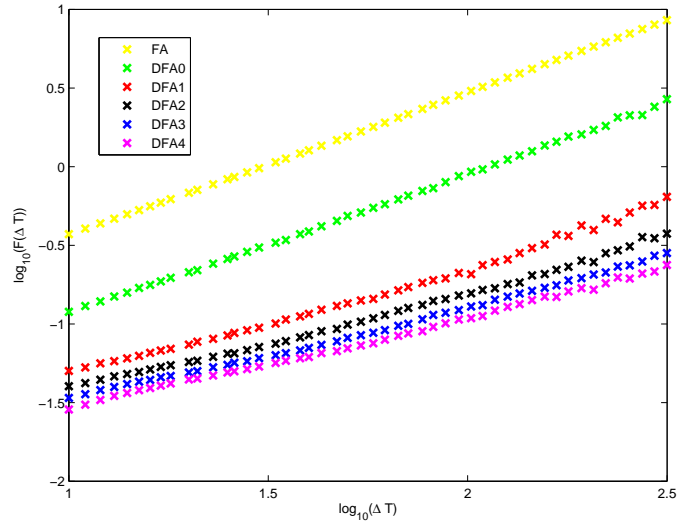


Figure 4.5: DFAn and FA of NASDAQ generalized volatility. Note: window size  $\Delta T$  is in days.

Figures 4.7 and 4.8 show results from DFAn and FA of N225 and FTSE indices respectively. It is important to note that DFA2 stands out as optimal fluctuation function estimator just as we discussed for DJIA index. The N225 index demonstrates increased LRC nature similar to DJIA and NASDAQ indices. The FTSE index data does not show predominant crossover phenomenon in fluctuation functions from any of the DFAn techniques applied.

Figure 4.9 shows the summary of all DFA plots by comparing DFA2 of all indices. We note that SSE and FTSE demonstrate distinguishable features from other indices. SSE has predominantly decreasing LRC nature and FTSE does not show crossover phenomenon. However, it is most interesting to note that the fluctuation function plots from DFA2 of DJIA, NASDAQ and N225 are close by each other and nearly indistinguishable. This shows the similarity between the three indices,

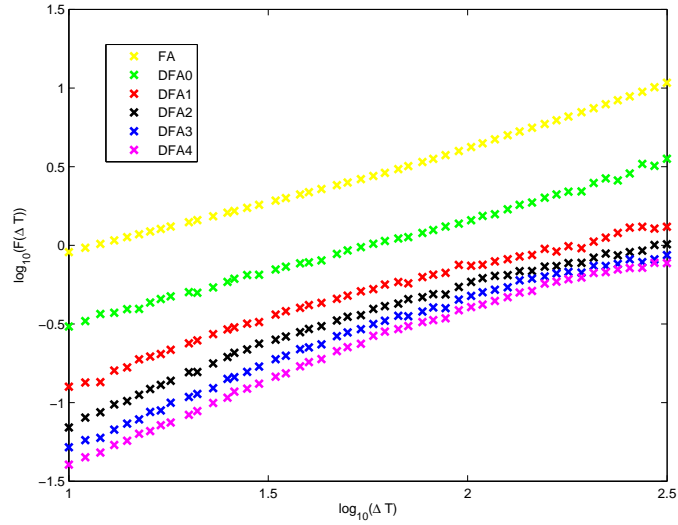


Figure 4.6: DFA $n$  and FA of SSE generalized volatility. Note: window size  $\Delta T$  is in days.

suggesting their interdependence with each other.

We made a case for scaling crossovers seen in fluctuation functions from DFA2 of various indices from the same period of time. By this analysis, we demonstrated similarities and differences in the crossover features of generalized volatilities of the indices. Now we are confronted with the question of what increasing and decreasing LRC nature means? In order to answer this better we first demonstrate two theoretical models for such varied nature of crossovers and discuss further what that means for the phenomenon seen in each index.

### 4.2.3 Convexity and concavity in LRC

We discussed briefly about increasing and decreasing LRC natures in Section 1.1. Those data that demonstrate increasing LRC nature through crossover

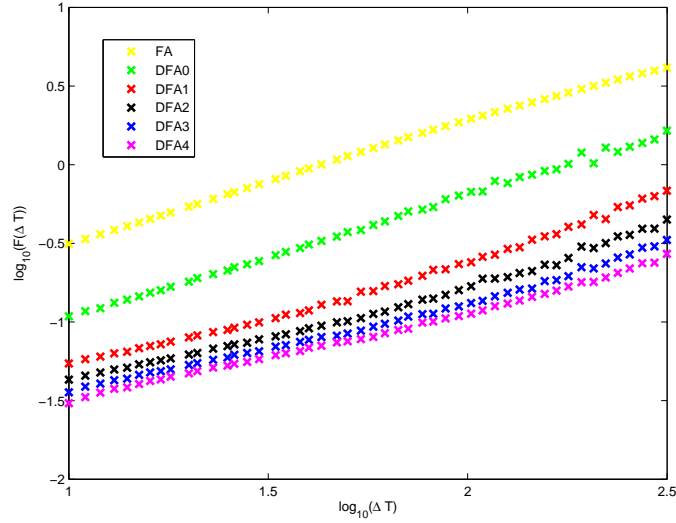


Figure 4.7: DFA $n$  and FA of N225 generalized volatility. Note: window size  $\Delta T$  is in days.

region are defined to be concave in crossover and decreasing LRC nature defined to be convex in crossover. There are significant differences between these two types of systems.

Systems with increasing LRC or concave crossover nature are typically modeled with additive noise and trend where the trend is non-stationary and the noise is stationary by definition. One such example is fractional Gaussian noise (fGn) added on top of fractional Brownian motion (fBm). These theoretical models were introduced in Section 2.2 earlier. The noise is modeled by fGn and the trend by fBm. This choice of trend is to assure single Hurst exponent with no crossovers in fluctuation functions, which are estimated as variance of detrended data. Variance of the signal with additive noise and trend components is also additive i.e.,  $F_{signal}^2(\Delta T) = F_{noise}^2(\Delta T) + F_{trend}^2(\Delta T)$ . For the DFA1 of fGn+fBm model shown

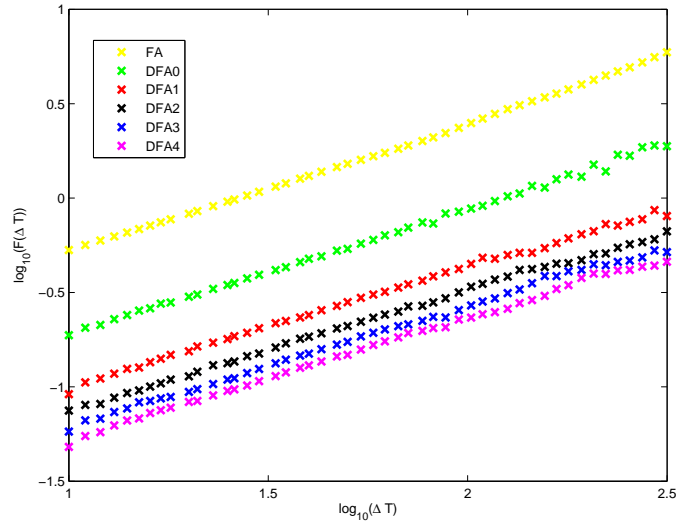


Figure 4.8: DFA $n$  and FA of FTSE generalized volatility. Note: window size  $\Delta T$  is in days.

in Figure 4.10, we used Gaussian noise ( $H = 0.5$ ) for fGn and its Brownian motion counterpart ( $\alpha = 1.5$ ). We can clearly see that the LRC increases through the crossover region or with increasing window lengths ( $\Delta T$ ).

Systems with decreasing LRC or convex crossover nature are typically modeled with noise integrated into trend (or dynamics). One such example is OU-Langevin. This theoretical models was introduced in Section 3.3 earlier. In this model, the noise represented by Gaussian noise ( $H = 0.5$ ) is added to the dynamics (or in the differential equation representing the model) thus making it a stochastic differential equation. For the OU-Langevin model we used in Figure 4.10, we used a decay parameter  $\lambda = 1/20$  and  $\sigma = 40$  to obtain the discrete time series after integration of Equation 3.1. Applying DFA1 to the time series, we can clearly see that the LRC decreases through the crossover region or with increasing window lengths ( $\Delta T$ ).

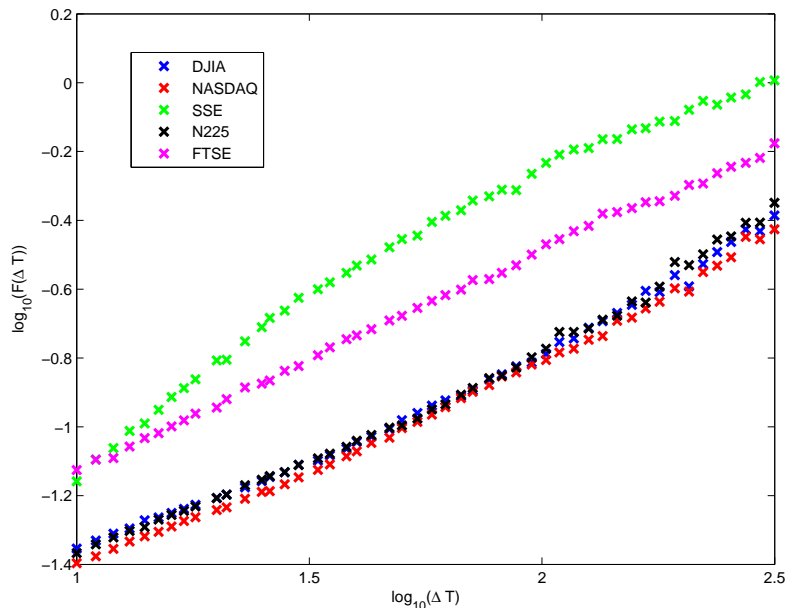


Figure 4.9: DFA2 of generalized volatility of all indices.

Now, with theoretical models presented and compared with each other, we would like to present the underlying difference between their natures. In the first model, representing concavity, the noise is added on top of the trend and in the second model the noise is added to the dynamics and subsequently integrated. This is a predominant difference to be noted because in one case the noise or the high frequency component is riding on top of the trend and in other case the noise is integrated into the process that causes the trend.

In Figure 4.11, we see the Hurst exponents from DFA2 of generalized volatilities of the indices we used for analysis before and after the crossover. The dotted line represents case where the Hurst exponent does not change through crossover or that there is not crossover. Concavity is exhibited by those indices above the dotted line and convexity by those indices below the right, respectively. Further the

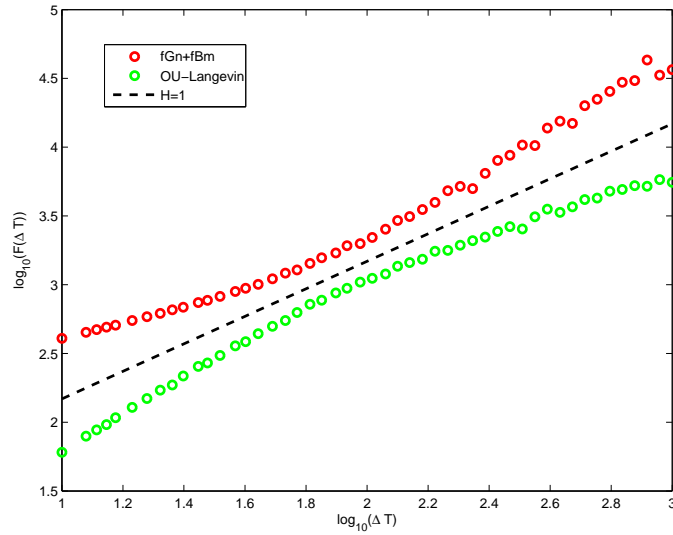


Figure 4.10: Comparison between concave and convex LRC seen in two theoretical models. Plots correspond to DFA1 of fGn+fBm and OU-Langevin models with a line representing  $H = 1$  shown for reference.

distance of a point from the line, greater is the crossover. This figure clearly shows the concavity and convexity in different indices.

From the theoretical models for concavity and convexity proposed earlier, it is important to note that in SSE which exhibits convexity, the noise within the volatility of the index are folded into the dynamics. This is in contrast to what happens in NASDAQ, DJIA and N225. In these three cases, the noise or speculative component is separable from trend. This implies that the market structure has separable low frequency (trend) and high frequency (noise) components in these three markets. In SSE, the speculations are closely integrated within the trends and their separation not as simple as in the three indices before. Further, FTSE does not have significant crossover or that the fluctuations and trends are almost of the same fractality or complexity.



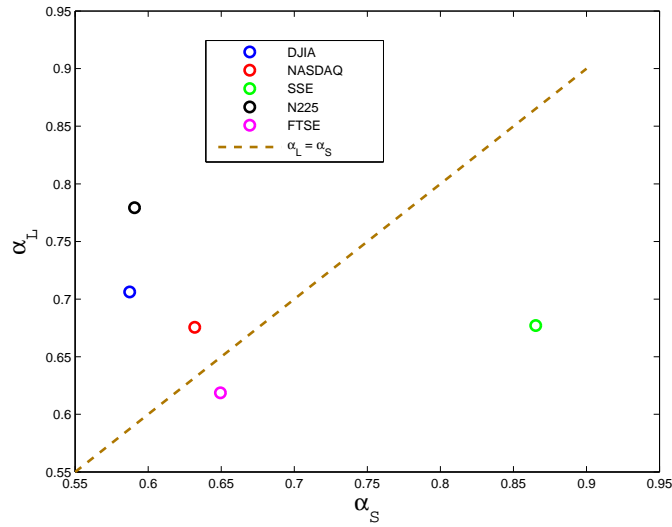


Figure 4.11: Concavity versus convexity seen in DFA2 of generalized volatility of various indices. The points show the separation between long time and short time DFA exponents ( $\alpha_L$  and  $\alpha_S$  respectively) in relation to dotted line that represents absence of crossover.

### 4.3 Conclusions

Results from our analysis establishes differences among different financial indices and shows how some markets are more informationally inefficient (from efficient market hypothesis) than others [Eom et al. 2008, Lillo and Farmer 2004]. What is even more interesting is the fact that all the markets demonstrate scaling crossover features in their generalized volatility data. Our primary goal was to extend the crossover analysis by Hyperbolic regression which was developed in earlier chapter for magnetosphere data to data from finance. Financial indices, particularly the generalized volatility in index data is known to show crossover in fluctuation functions. Previous work has shown the existence of such crossovers but has fallen short

of a rigorous analysis. We pick up the idea and extend our crossover analysis to generalized volatility of five largest global financial indices.

From the crossover analysis, we see that there are differing features with respect to increasing or decreasing LRC natures within the data. In order to explain the underlying cause of such differing LRC natures, we presented theoretical models. Concavity in crossover is seen in systems where noise is separable from trend and convexity is seen in systems where noise is integrated into dynamics of the system. Using such models we could explain the concave and convex natures seen in financial data on a global scale. There are aspects in these results that needs deeper understanding of economics and finance to better explain them. It is important to note that we demonstrate an application of our regression scheme to reveal unique and illuminating aspects allowing us to understand multi-scales features in financial data better. This work can be considered to be a first step for further analysis of concave versus convex nature of crossover phenomenon in financial data and utilizing such features to understand the nature of markets better.

## Chapter 5: Scaling crossover in heart rate variability : Extent of crossover analysis

### 5.1 Overview

We applied crossover analysis utilizing hyperbolic regression to data from space weather in chapter 3 and financial data in chapter 4 respectively. From the analysis on space weather data, we could estimate important multi-scale parameters such as the Hurst exponents before and after crossover and the crossover time scale. The crossover time scale is a measure of the time scale of internal dynamics of such driven systems. From our analysis on the data from finance in the last chapter, we have shown differences in nature of crossovers seen in composite indices on a global scale. Furthermore, using theoretical models, we could see the interdependence between the high frequency speculative component that changes on a much faster scale and the trend. Not all systems are equal with respect to separability between the two components as we have shown. This has been the central theme of work in chapter 4.

Crossover in fluctuation functions from heart rate variability (HRV) data was identified earlier [Peng et al. 1993; 1995; 1998]. The existence of crossovers result in

alteration of the scale free nature of the fluctuation functions, typically estimated by Detrended Fluctuation Analysis (DFA) due to non-stationary nature of such data. An introduction to HRV data was provided in Section 1.2. Existence of long range correlation (LRC) phenomenon in physiological processes [Stanley et al. 1994] necessitates redefining homeostasis taking into account such ideas [Goldberger 1991]. The essential element to be noted here is that the pacemaker cells in heart are controlled by competing processes which account for homeostasis (due to sympathetic) and LRC (due to parasympathetic) natures respectively. This aspect is discussed in further detail in Section 5.2.4 where we attempt to interpret our results from a dynamical theory viewpoint. Existence of a single crossover in fluctuation functions is a common feature between HRV and AL index data (as seen previously in chapter 3). In this chapter, application of hyperbolic regression on HRV data from patients with congestive heart failure (CHF) and those with normal sinus rhythm (NSR) shows some interesting discerning features between the two cases. Patients with undiagnosed CHF are prone to a severe risk of a heart stroke and potentially death. A parameter to measure the extent of change in scales, labeled as extent of crossover parameter, is introduced and used to quantify the difference between the two cases.

There have been previous efforts to model HRV data using approaches from chaos theory [Goldberger 1996, Goldberger et al. 1990, Shelhamer 2006]. Given the complexity in underlying process leading to HRV data, it was assumed that such modeling would be valid. Poincare plots was commonly used to analyze HRV data. Approximate entropy and correlation dimension were used to describe the complex-

ity of HRV data [Storella et al. 1994]. However, lack of evidence for low-dimensional chaotic nature of HRV data has been shown through further analysis [Kanters et al. 1994]. Hence, it is important to note that in our current work, we do not attempt to model HRV data using dynamical features. However, using purely data derived approaches, perform statistical analysis to uncover features that can potentially help in detecting ailments. As a first step, we present evidence of scaling crossover seen in DFA of HRV data in the following section.

## 5.2 Scaling crossover in HRV

Hurst exponent ( $H$ ) is a measure of LRC in time series data. Rescale range (R/S) analysis was the first method proposed to measure H exponent, followed by fluctuation analysis (FA). FA works well for stationary data, and fails for non-stationary data with an inbuilt trend in them. These techniques were introduced in Section 1.3. DFA overcomes this challenge by addressing typical additive trends of polynomial orders and estimates an exponent,  $\alpha$  with known relationship to H as  $\alpha = H$ , for stationary data, and,  $\alpha = H + 1$ , for non-stationary data, even in the presence of such trends. In estimating  $\alpha$ , the method assumes that the fluctuation functions estimated as the variance of detrended random walks are scale free with respect to the length of the time window of estimation. We shall label as DFAn, the DFA scheme when applied to a system with an  $n^{th}$  order polynomial used for detrending.

We obtained cardiac interbeat interval data from physnet ATM service (<http://physionet.org/>

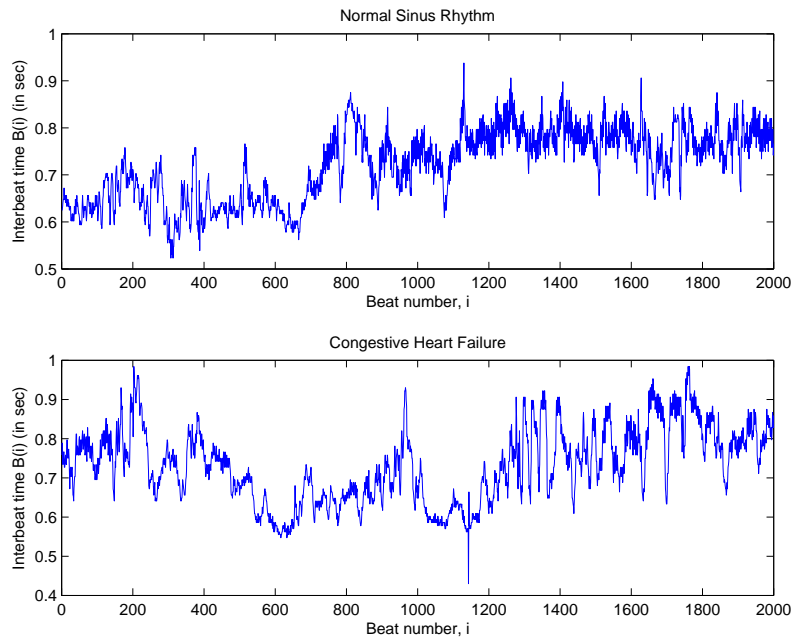


Figure 5.1: Interbeat interval data for 2000 beats of normal healthy subject (upper panel), and, patient with CHF (lower panel).

[bin/atm/ATM](#)) for patients with normal sinus rhythm (NSR) and those with congestive heart failure (CHF). The RR intervals were extracted from ECG data of 15 patients from each group. Such an example of interbeat interval time series is shown in Figure 5.1 for a patient each with NSR and CHF conditions respectively.

### 5.2.1 Crossover in RR interval data and previous results

Previous work [Peng et al. 1998] that applies DFA to RR interval data has shown existence of time scales with different Hurst exponents. Results from this analysis are shown in Figure 5.2. The fluctuation analysis using DFA technique was performed on the interbeat times with respect to windows lengths of beat numbers. It can be seen that NSR case and CHF case are opposite in their LRC nature at

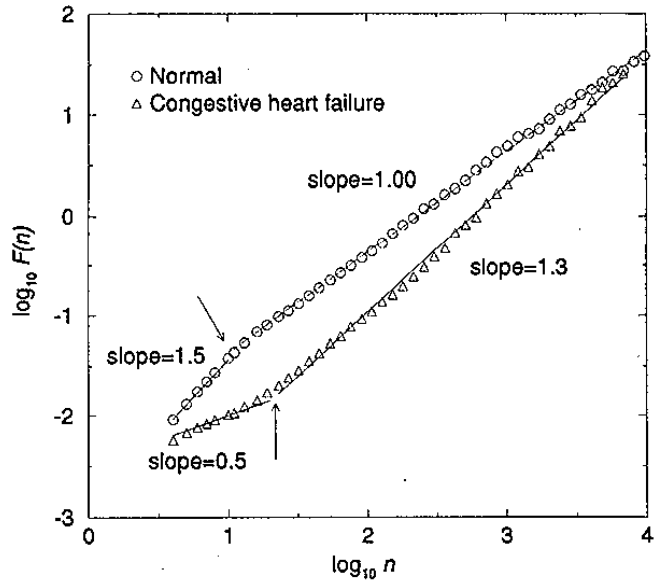


Figure 5.2: Result from previous work [Peng et al. 1998] showing DFA of RR interval data from normal (NSR) and congestive heart failure (CHF) cases. We see that the NSR case has a decreasing LRC while CHF case has increasing LRC with time respectively. The window lengths here are in units of beat numbers.

long terms. The LRC nature in NSR case is decreasing, indicative of dominance by sympathetic nervous system at long times leading to homeostasis. The opposite is true in CHF case where increasing LRC with time indicates dominance by the parasympathetic nervous system.

Using the new regression scheme described in section 3.2, we now study not only the scales before and after the crossover, but also quantify the extent of crossover or change in scales that occur in such data. This helps us compare the exponents previously obtained using a manual scheme to select the point of crossover and draw new conclusions by studying the relationship between the extent of change

in scales and the nature of cardiac health of a patient.

We measure the extent of the crossover (EOC) parameter ( $\theta_c$ ) as the separation of angle between the two asymptotes:

$$\theta_c = \arctan(\alpha_l) - \arctan(\alpha_s) \quad (5.1)$$

where,  $\alpha_l$  and  $\alpha_s$  are the long term and short term DFA exponents respectively. EOC (measured in degrees( $^\circ$ )) would define the extent and the nature of the crossover. As an example, if  $\theta_c$  is positive, it implies extent of LRC in the system is increasing, and, when  $\theta_c$  is negative, LRC in the system is decreasing.

### 5.2.2 Crossover analysis for NSR cases

Previous work [Peng et al. 1998] claims that NSR cases have a negative  $\theta_c$ , and, the crossover occurs at about 10 beats. Our analysis utilizing hyperbolic regression is not in agreement with this conclusion as crossovers in the system at lower beat numbers are dominated by crossovers at higher beat numbers. Existence of multiple crossovers complicates the analysis and the resulting hyperbolic fit would be equivalent to fit using linear regression in such cases. The analysis can be seen in Figure 5.3 for one of the cases in our group of 15 subjects with normal sinus rhythm. Existence of multiple crossovers throws off course the hyperbolic regression scheme and in essence we obtain a linear fit with parallel asymptotes in many cases.



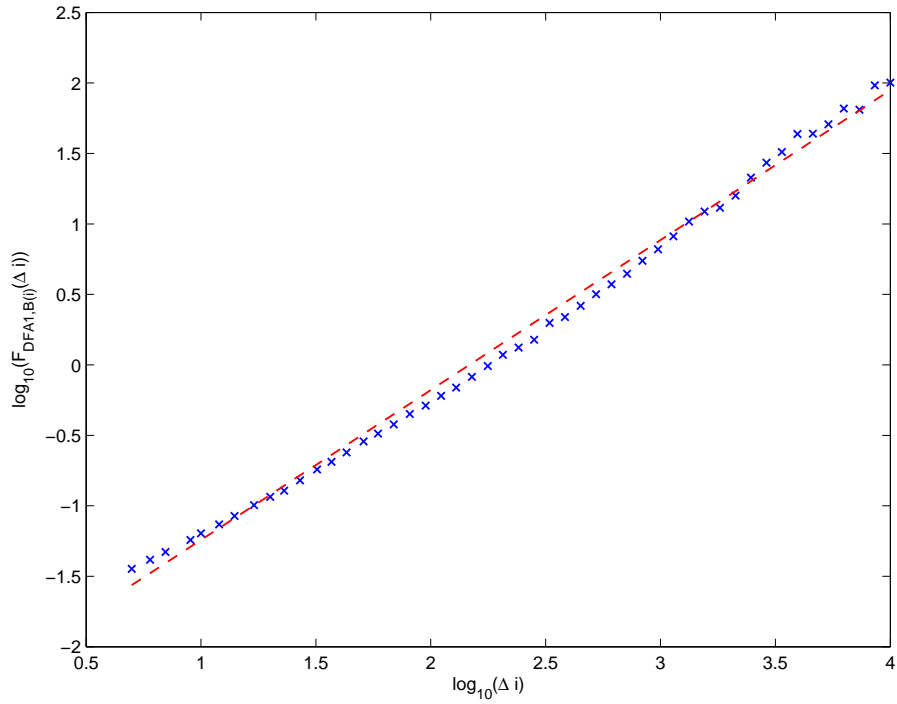


Figure 5.3: Hyperbolic regression and asymptotic fit of fluctuation functions from DFA1 of RR interval data of a NSR subject with DFA1 in blue and regression fit in red. The regression identifies that the two asymptotes are parallel or that the linear fit is the best fit.

### 5.2.3 Crossover analysis for CHF cases

The hyperbolic regression scheme is applied to the fluctuation functions from DFA1 of RR inter-beat interval data for 15 CHF patients. Crossover analysis on DFA1 of a CHF patient can be seen in Figure 5.4. EOC is measured as  $\theta_c = 22.31^\circ$  for this particular case. In the next section we discuss in detail the interpretation of extent of crossover presenting an explanation from dynamical theory viewpoint of differences in EOC between NSR and CHF cases.

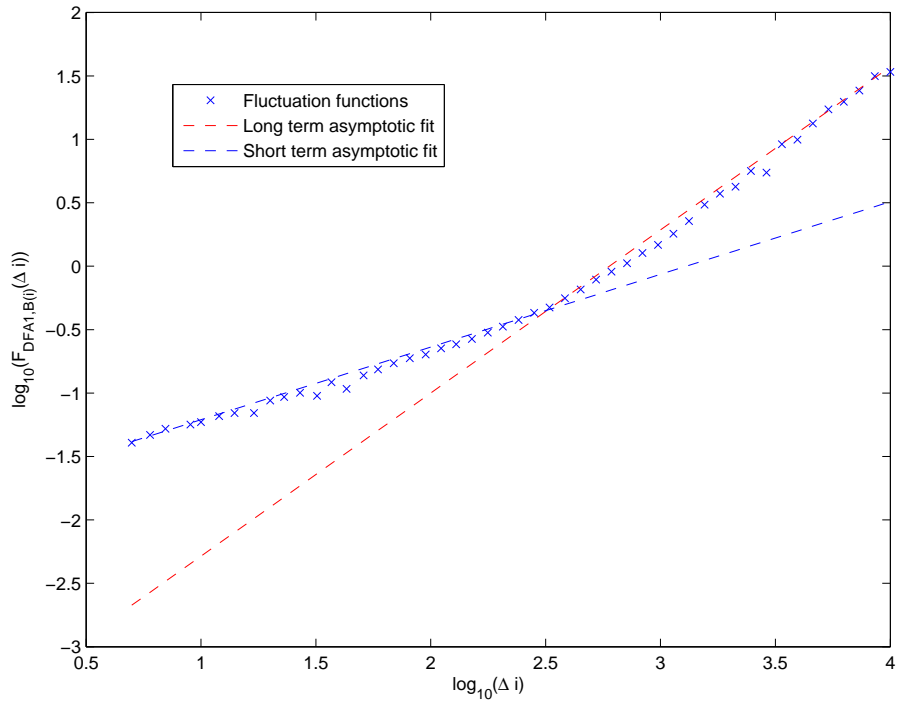


Figure 5.4: Hyperbolic regression and asymptotic fit of fluctuation functions from DFA1 of RR interval data of a CHF patient. The angle between the asymptotes (red and blue dashed lines) in this particular case is measured to be  $\theta_c = 22.31^\circ$ .

#### 5.2.4 Extent of crossover analysis

As discussed earlier in Section 1.2, sympathetic and parasympathetic nervous systems actively compete for control over pacemaker cells [Levy 1971]. The sympathetic nervous system is responsible for flight or fight response and the parasympathetic for breed and feed response. Due to instinctive nature of sympathetic nervous system, it causes sudden changes in the performance or rate of heart beat and hence is the driver of high frequency component of HRV data. Parasympathetic acts over extended time scales and does not show up as in the high frequency. The LRC

seen in HRV data is mainly due to parasympathetic nervous system and the lack of correlations due to sympathetic. This is considered to be the dynamical basis for function of pacemaker cells.

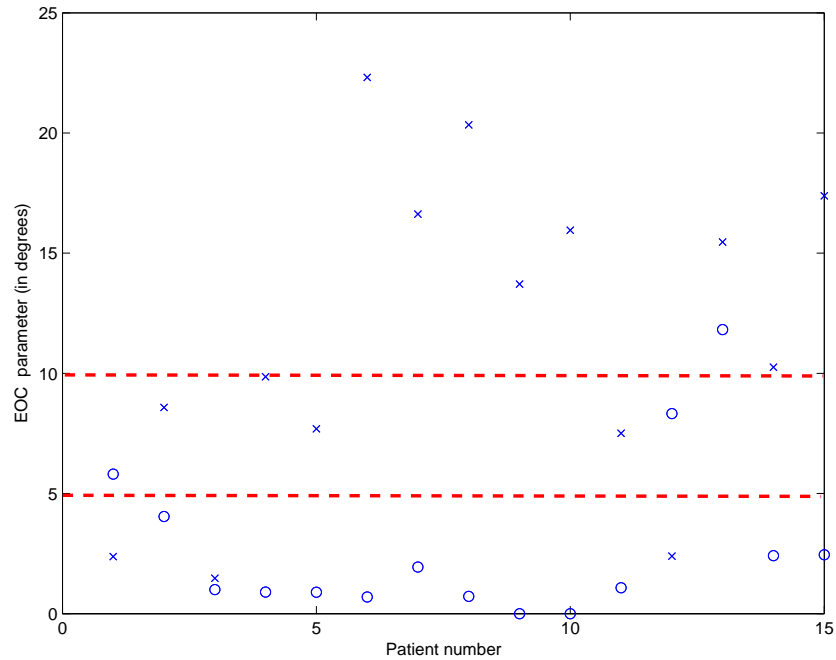


Figure 5.5: EOC for NSR and CHF patients. The EOC parameter for NSR is shown in circles and for CHF in crosses. We see that  $\theta_c$  for most NSR patients is below  $5^\circ$  and for most CHF patients is at or above  $10^\circ$ .

From results in Figure 5.5, it is important to note that 8 of the 15 cases had values less than  $1^\circ$  indicative of very small separation in scaling exponents before and after the crossover in majority of NSR cases. For a system to change scales between  $\alpha_s = 0.5$  and  $\alpha_l = 1.5$  asymptotically as postulated earlier [Peng et al. 1998],  $\theta_c$  would have to be  $29.74^\circ$ , and, none of the patients from our group show such a large value of the parameter.

However, interestingly 8 out of the 15 CHF patients have values greater than

$10^\circ$  for EOC. The minimum value from this group is  $1.5^\circ$  and maximum value is  $22.31^\circ$ , and, it must be noted that all are positive indicating consistently increasing LRC pattern. Hence,  $10^\circ$  can be used as the cutoff value to differentiate between cases with CHF and NSR. In total, out of the 15 CHF cases: 8 cases ( $\approx 50\%$ ) had EOC above  $10^\circ$ , 4 cases between  $5^\circ$  and  $10^\circ$  and 3 cases ( $20\%$ ) had values less than  $5^\circ$ . Hence, cases with  $\theta_c$  between  $5^\circ$  and  $10^\circ$  can be treated as border-line cases.

Also, for the 15 NSR cases: only 3 cases ( $20\%$ ) had EOC greater than  $5^\circ$ , with just 1 case having a value slightly greater than  $10^\circ$ . The rest of the cases i.e., 12 cases ( $80\%$ ) had values less than  $5^\circ$ , and much close to  $0^\circ$ .

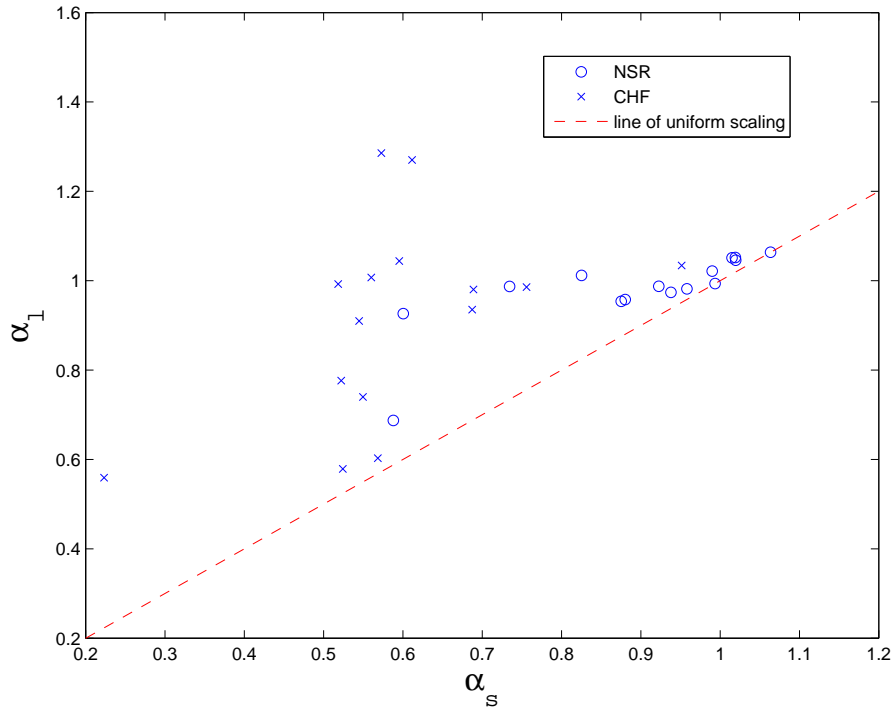


Figure 5.6: Scatter plot showing  $\alpha_s$  and  $\alpha_l$  for short and long terms respectively for NSR and CHF patients with dashed line representing  $\alpha_s = \alpha_l$ . We see that the  $\alpha_s$  is very close to  $\alpha_l$  for most NSR patients and the difference in CHF patients is apparent.

For the cases we analyzed, the scaling exponents before the crossover were measured as:  $\alpha_s = 0.59 \pm 0.15$ , and,  $\alpha_l = 0.9135 \pm 0.22$  in CHF cases. This is indicative of a transition from correlationless Gaussian noise like nature in short term to highly correlated Gaussian noise/anti-correlated Brownian motion like nature in the long term. Results for the comparison between the exponents before and after crossover as measured by hyperbolic regression scheme in NSR and CHF cases can be seen in Figure 5.6. The greatest value of  $\alpha_l$  from this group is 1.29 and there were only 5 cases out of 15 cases with values greater than 1 indicating that the previous claim of  $\alpha_l$  1.5 in the long term may not be totally valid, or, at least unverified from the analysis we performed on our group. The extent of crossover analysis as seen in Figure 5.5 essentially reflects in the results we see in Figure 5.6.

Measurement of crossover times from DFA1 may not be a valid proposition due to the detrending operation, but can give an idea of correlation times in data or the sampling time required to transit between regions of differing correlation properties in the system. From our group, the crossover times are found to be:  $T_c = 303 \pm 135$  seconds or in terms of the number beats,  $i_c = 452 \pm 192$  beats.

From this analysis, we have shown the difference in EOC between NSR and CHF cases. Now, we present a discussion covering the interpretation of these results. As mentioned earlier in this section, sympathetic and parasympathetic nervous systems try to compete for control over the pacemaker cells in the heart that controls HRV. In patients with NSR, there is an equilibrium that is established between the two nervous systems. However, existence of LRC implies departure from correlationless system. Due to better health of heart in NSR, external pressures requiring

increased function of heart does not significantly disturb equilibrium between the two autonomous nervous systems. Hence, there is no significant crossover phenomenon exhibited by HRV data in these patients.

CHF patients usually fail to provide enough blood in conditions where external influences necessitate increased heart function [McDonagh et al. 2011]. This increased requirement in blood flow forces the heart to go out of equilibrium where the sympathetic and parasympathetic nervous systems fail to reconcile with each other. Due to this, the two components become separable and seen causing crossover phenomenon in fluctuation functions. The concavity seen through EOC estimation is a direct proof of increasing LRC. This can be considered to be the reason why we see predominant crossover phenomenon in CHF as compared to NSR cases.

### 5.3 Conclusions

We found that results of our analysis on RR interval data agree with previous work on many counts, most important of those is the increasing LRC in CHF cases data due to the violation of homeostasis. We substantiate the same by a parametric measure, the extent of crossover parameter, which estimates in degrees of angle the amount of change in the scaling exponent before and after the crossover. The parameter helps us see major difference in the nature of evolution of LRC in the data between NSR and CHF cases. Hence, we propose using it as estimated from 24 hour ECG measurements of patients as an early indicator of CHF. Clearly the results from 15 cases of each group and the statistics of the results are promising

for the proposal to use EOC parameter as a diagnostic metric.

However, it is important to note that more work to benchmark this parameter is required which requires compiling RR interval data from many more patients of NSR and CHF groups. Also, further analysis is necessary to correlate EOC parameter with the severity of CHF in patients. We hypothesize that increased EOC indicates increased levels of the CHF ailment. This would imply that one can directly use the parameter as a measure of the severity of the disease, which also requires further work to establish this correlation. We have presented crossover analysis on data from three systems that are completely different from each other—space weather, finance and physiology. Each case presents a unique challenge with respect to outcome of our analysis and we addressed the challenge with specific analysis in each of the cases.

Analysis on AL data from space weather has shown the importance of measuring crossover time. Due to our inability to measure the time scale from DFA, we presented an alternative approach utilizing modeling it using a stochastic system that embodies similar crossover phenomenon. Application of the crossover analysis on data from finance has shown the differences between the nature of markets in terms of concavity versus convexity in crossover phenomenon. Crossover analysis on HRV data from physiology in this chapter has allowed us to see the predominant differences between two class of patients—those with normal heart rhythm (NSR) and those with an ailment (CHF). Using a proposed metric, we can quantify the difference between these two cases.

## Chapter 6: Characterizing predictability using fluctuation analysis

### 6.1 Overview

We presented examples of complex systems that exhibit evidence of multi-scale nature via scaling crossover seen in fluctuation functions in chapters 3, 4 and 5. Another application of fluctuation analysis has been to use the Hurst exponent of the high frequency component or noise in the system. This exponent can be related to a measure of possibility of extreme event occurrence or the tail exponent of its probability distribution. This is central to the approach taken in the work of this chapter.

Characterization of predictability of data generally refers to study of residuals after predicted values are removed from actual values. This usually involves estimation of a mean square value of the residuals as shown in section 1.5 which is widely used as an estimate of accuracy of the prediction. However, we tend to ignore the fact that errors, although seemingly normally distributed, may incorporate heavy tails in them. These are a direct result of our inability to predict or a deterioration in the quality of prediction. Conventional characterization does not factor in this possible heavy tailed distribution of resulting error terms.

Study of a heavy tailed distribution usually involves the estimation of a tail



exponent. The fatter the tail, the larger the value of the exponent, by definition. However, the estimation of this exponent from data is highly prone to biases in data, making its estimation very unreliable. In light of this realization, recent studies show that the degree of self similarity in data, reflects in the heavy-tailedness of its distribution [Taqqu et al. 1995]. This dependence helped empirically relate the Hurst exponent, a measure of self similarity in data, to the tail exponent [Stoev et al. 2011].

Estimation of the Hurst exponent from a data requires detrending to remove trends. This process is often complicated if done by *a priori* assumption of the nature of the trends as done in the DFA technique [Hu et al. 2001]. Time series predictions based on dynamics of a system often represent the trends well, as trends contribute to dynamics of a system. This allows us to detrend the data using the predicted values from historical time series for a data. We label this method as Fluctuation Analysis after Trend Elimination (FATE), which can be a broad class of techniques that allow one to remove trends *a priori* before estimating the Hurst exponent.

Nonlinear time series prediction techniques are widely used to predict AL data [Chen and Sharma 2006, Ukhorskiy et al. 2002, Valdivia et al. 1996]. This involves a series of steps as described in Section 1.5, starting with attractor reconstruction and then using any of the filters to use information from the reconstructed phase space to predict future time step(s).

We begin by showing an example of this design as applied to a well known theoretical model in section 6.2. In our study, we illustrate the working of FATE

technique on data from finance (Microsoft daily closing stock prices), and, the AL-Solar wind system as seen in section 6.3. As a comparison, we show the tail exponents as estimated by another well established technique i.e., the method of Hill estimator to relate the tail exponent estimated using Hurst exponent from the tail exponent given by the data, in section 6.4, followed by conclusions in section 6.5.

## 6.2 Technique demonstrated on a theoretical model

We demonstrate the working of FATE technique, and, use theoretical model of correlated noise given by fractional Gaussian noise (fGn) with  $H = 0.9$  on a sinusoidal trend (of the form  $A \sin(\omega t)$ ). The value of amplitude of trend is chosen in relation to the standard deviation of fGn. This is to assure that the crossover in fluctuation functions is apparent and the trend does not dominate over the noise.

Time delay embedding technique is used to reconstruct the phase space as a trajectory matrix to be used as training set to search for nearest neighbors to predict the dynamics of the system through a local linear weighted mean field filter. Details of this method are described in detail in Section 1.5. This method of prediction helps us remove the sinusoidal trend near completely and thus the resulting detrended data shows  $H = 0.83 \pm 0.02$ , very close to the  $H = 0.86 \pm 0.01$  of the fGn used. We apply DFA2 to estimate  $H$  exponent as FA fails in the case of data with sinusoidal trends to estimate Hurst exponent. These results can be seen in Figure 6.1.

After application of FATE, we retrieve the right Hurst exponent of the underlying variability as modeled using fGn data. The  $H_{FATE} = 0.83 \approx 0.86$  of the

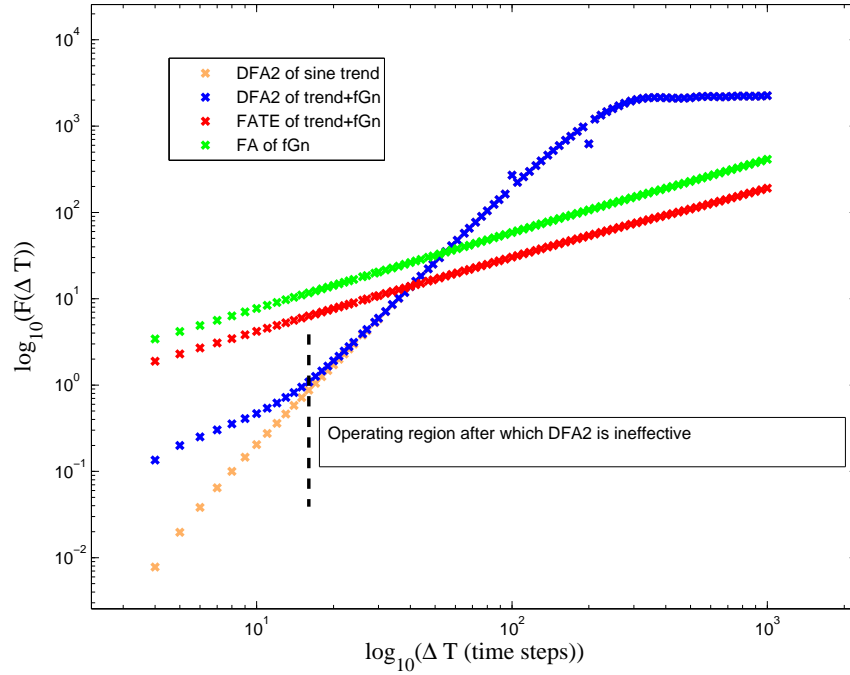


Figure 6.1: FATE applied to test data with fGn ( $H = 0.86$ ) riding on top of sinusoidal trend. We see that FA of detrended data retrieves most of the scaling properties of the fGn with explicit detrending of sine trend using time series predictions.

actual data.

Open complex systems often show heavy tailed distributions. This is reflected in the power law dependence of probability distribution function (pdf) of a time series as:  $pdf(x) \propto x^{-\gamma}$  as  $x \rightarrow \infty$  where  $x$  is the magnitude of an event from within the time series. Taqqu's theorem [Leland et al. 1994, Taqqu et al. 1997] relates the tail exponent  $\gamma$  (a measure of heavy-tailness in probability distributions) to the Hurst exponent  $H$  (a measure of extent of LRC) as:

$$H = \frac{3 - \gamma}{2} \quad (6.1)$$

From this result and the  $H_{FATE}$  we obtained, the expected Hurst exponent is  $\gamma = 1.33$ .

### 6.3 Tail exponent estimation using Hill estimator

Hill estimator technique provides a way to estimate the tail exponent from data using construction of a series that converges to the exponent [Hill 1975]. It must be noted that the technique is computationally efficient and easy to be implemented on large data sets. Using Hill estimator technique, we can estimate the tail exponent of the fGn data used in previous section to be 1.29, which in turn corresponds to a  $H$  of 0.855 very close to the actual Hurst exponent of the data used (0.86).

This is a direct illustration of the FATE technique and the ensuing result which accurately estimates the tail exponent of the data. It must be noted that this method does not directly estimate the tail exponent, the process of which is known to be riddled with data size biases and scaling crossovers in histograms. Such issues result in what are referred to as Hill horror plots in the technique of Hill estimator [Resnick 1997, Resnick and Starica 1997].

### 6.4 FATE and characterization of tail exponent

We apply the demonstrated method to our data from AL index and the Microsoft stock price (daily closing values).

### 6.4.1 AL index data

The technique of nonlinear time series prediction of the AL and the Dst indices is well established and has been extensively used to forecast space weather. Method of prediction is described in greater detail in Section 1.5. Reconstruction of phase space requires historical time series values. Typically, more data used yields greater detail of the phase space, thus improving the quality of prediction. Currently, a routine that forecasts space weather by using real time solar wind measurements is active and results are shown on our website: <http://spp.astro.umd.edu/gmprediction/>.

The data used for phase space reconstruction was prepared using 1 minute AL values, and,  $v$ ,  $B_z$  values from the ACE explorer measurements. Links to these sources are given in Section 1.2. Data taken from respective sources for the years 2003-2013 was averaged into 5 minute values to standardize the sampling frequencies of all three data variables. Following the preparation of historical data for phase space reconstruction, we use the trajectory matrix obtained as the search space to find the nearest neighbors, thus allowing us to predict the next step. The details of this approach are described in Section 1.5. In the process of phase space reconstruction, one can model the system as independent by itself or as one being driven by another system i.e., input-output. We label these two methods of modeling as autonomous and combined models respectively. Additionally, a model which does not use the data of the magnetosphere i.e., the input is also shown, which we shall refer to as non-autonomous.

Results for prediction from the three models are shown in Figure 6.2 for 2000

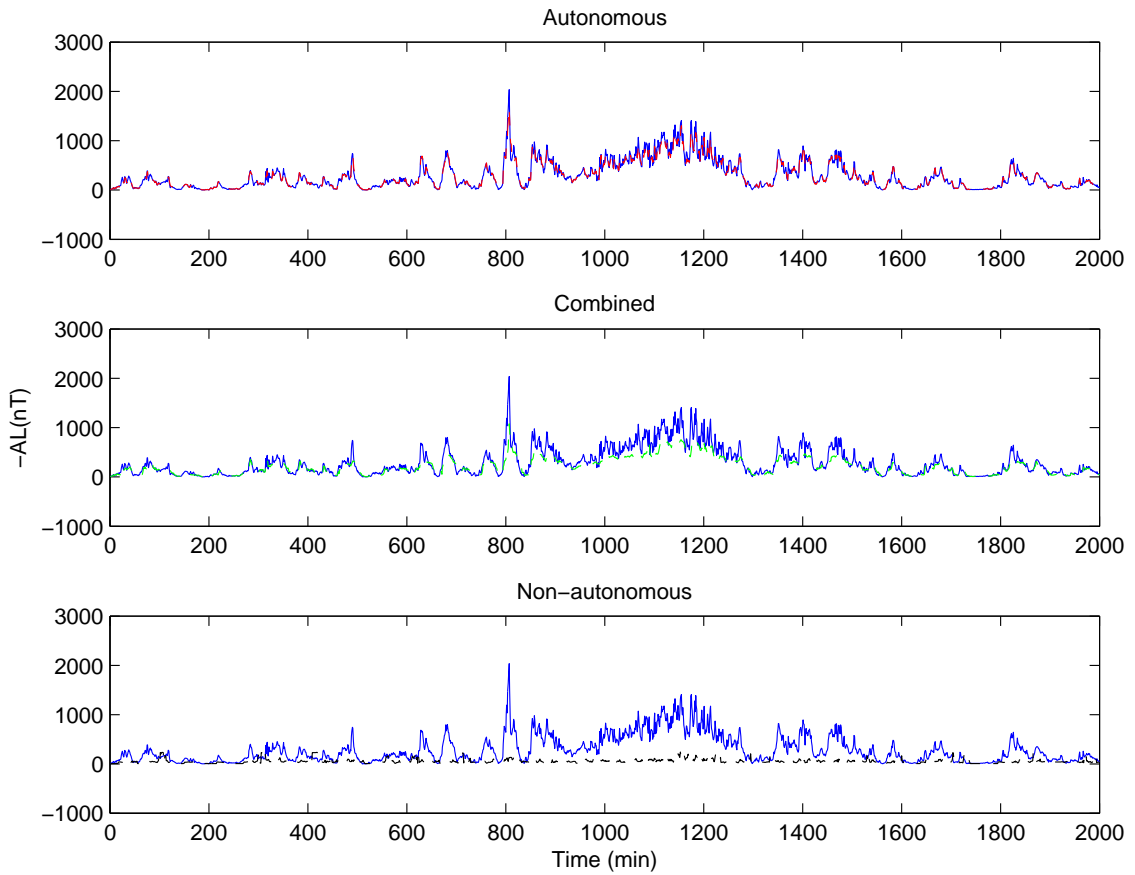


Figure 6.2: AL (actual in blue) and predicted (in red, black and green) for autonomous, combined and non-autonomous methods of predictions

time steps. Predictions were performed for data corresponding to 2012-2013 utilizing trajectory matrix obtained by historical data from 2000-2012. The phase space is reconstructed with a time delay of 1 time step (= 5 minutes) and an embedding dimension of 12. For the weighted mean field filter applied for predictions, the number of nearest neighbors was taken to be 10.

The results for the application of FATE on AL data can be seen in Figure 6.3. As expected, the non-autonomous model results in least difference from the fluctuation properties of the original data as seen in black and blue plots respectively.

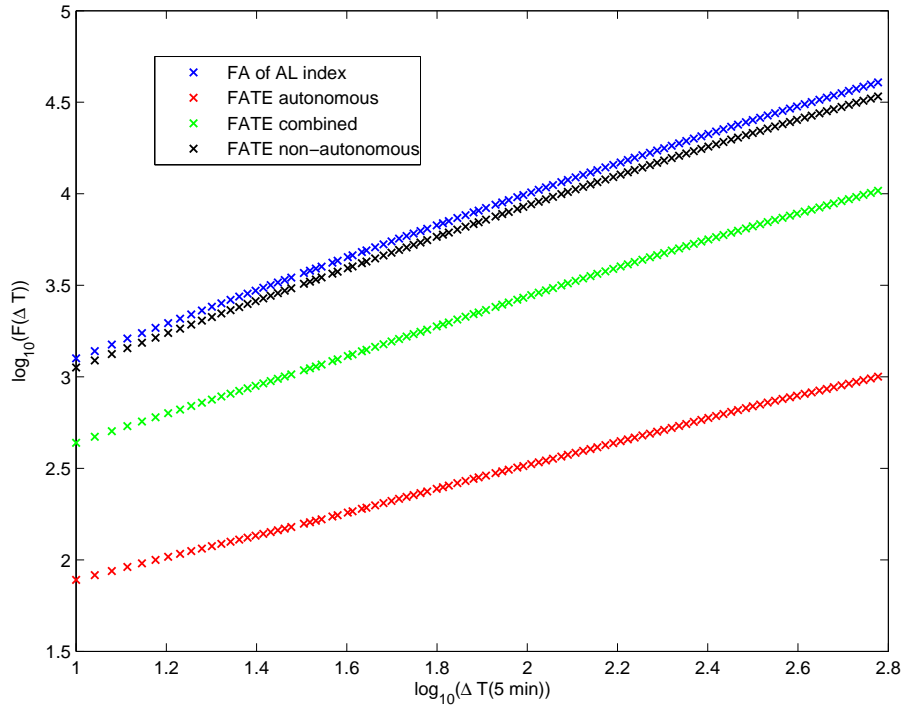


Figure 6.3: FATE applied to AL data with several prediction methods. We see that non-autonomous is not very different from FA of AL data. Also, autonomous predictions result in  $H = 0.5$  with most of crossover effects removed.

The Hurst exponent of the AL index not accounting for any crossover effects is  $0.84 \pm 0.01$  (slope of blue plot), and, the Hurst exponent of the FATE applied to AL with non-autonomous modeling is  $0.83 \pm 0.01$ . This minimal effect of removing the trend with non-autonomous model for predictions can be seen directly in the negligible difference in the Hurst exponents.

The Hurst exponent of the FATE applied to AL with autonomous modeling is  $0.63 \pm 0.01$  (slope of red plot), and, with combined modeling is  $0.78 \pm 0.01$  (slope of green plot). It must be noted that relatively low standard errors in all the Hurst exponents may be due to the increased averaging of fluctuation functions from

fluctuation analysis technique used for estimation in all these models. Clearly, the method of autonomous modeling detrends the data best of all periodic effects in the original data, and yields H exponent closest to 0.5, corresponding to uncorrelated Gaussian noise. As expected, the combined model still retains effects of trends and yields an intermediate value of Hurst exponent between that of actual data (trend+noise) and the noise.

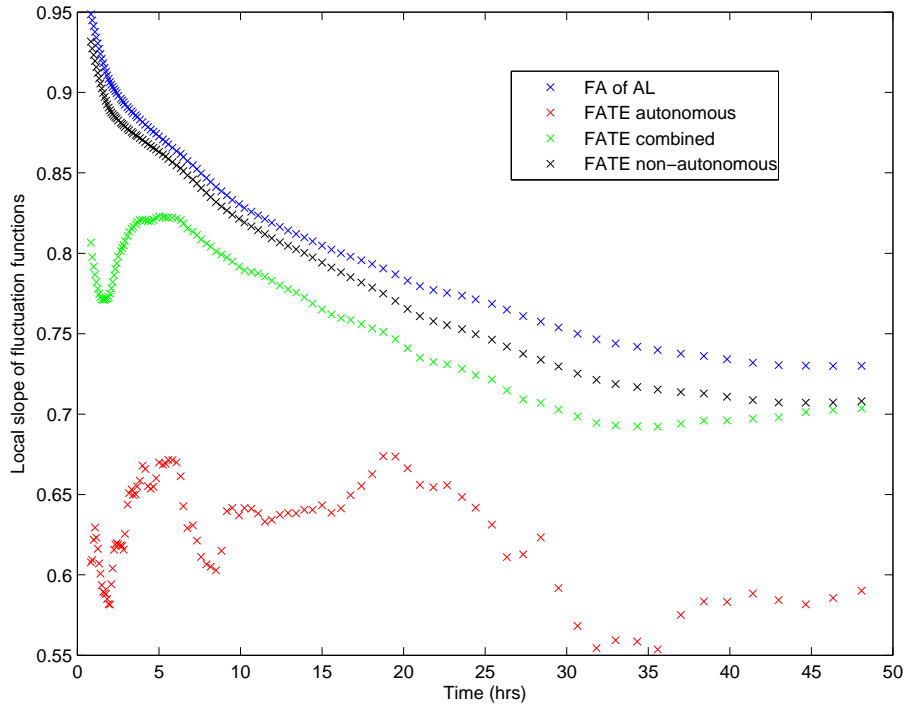


Figure 6.4: Local slope of FATE on AL. We see that the non-autonomous model (in black) retains most of the crossover effects shown in AL data (in blue).

The ability of each prediction model in detrending the system has been demonstrated from the results in Figure 6.3. We have seen in Chapter 3 the existence of crossovers in fluctuation functions of AL index data. Local slopes of the fluctuation



functions from Figure 6.3 are estimated as the slope of the power law fit between two consecutive time windows. This local slope picture as shown in Figure 6.4 helps us understand the point of crossover and the exponents before and after better by capturing changes in the exponents. Also, local slopes from each of the models shown in Figure 6.4 shows the ability of the respective models to remove the crossover effects in fluctuation functions. The local slopes in fluctuation functions from each model and the original data show an asymptotic approach to a long term exponent value after 24 hours which is much greater than the internal dynamics time scale of 4 hours 25 minutes as estimated previously in chapter 3. The local slope picture shows us that this crossover remains in the FATE that uses non-autonomous modeling and to an extent in the one that uses combined modeling. However, FATE with autonomous modeling best addresses this crossover and in essence removes the crossover nearly completely resulting in constant local slopes.

Errors are usually expressed in terms of the standard deviation of residuals. One such calculation, known as normalized mean square error (NMSE) was discussed in Section 1.5. The NMSE values for autonomous, combined and non-autonomous models are 0.1126, 0.4270, and, 1.0172 respectively. Effect of each of these method on the quality of predicted AL values can be further characterized using the resulting Hurst exponents. The basis for this approach, as described on the test case, is that if one were to predict the system well then the residuals or the error terms would be void of correlations. Or, the Hurst exponent of the errors would be close to 0.5-of that of correlation less Gaussian noise. To correlate the estimated exponents with the tail exponents of the respective errors, we estimate the tail exponent using

Hill estimator as described in Section 6.3. Using the Hurst exponents estimated by FATE, one can estimate the expected tail exponent using Taqqu's theorem. For the autonomous model of prediction, the expected tail exponent is  $\gamma = 1.74$ , and the tail exponent from Hill estimator is  $\gamma = 1.78$ . Combined model yields expected tail exponent as 1.44, where as the tail exponent using Hill estimator is 1.48. The difference in the Hill estimate from different models is due to the fact that errors in prediction differ in their heavy-tailness (as measured by their respective tail exponents) for predictions made using different modeling assumption i.e., autonomous or combined in this case.

We find that the non-autonomous model and the actual data still retain the scaling crossover effects in their respective fluctuation plots as seen in Figure 6.4. Existence of multi-scales as evidenced by the crossover makes the estimation of a single tail estimation complicated and thus unreliable. The Hurst exponent estimated by FATE using non-autonomous model of 0.83 implies a tail exponent of 1.34. However, the Hill estimator technique is non-convergent and make estimation of a single tail exponent unreliable.

These results on AL index data show that we can use nonlinear time series predictions as a measure of trend. This dynamical trend, when used to detrend the data before hand, allows us to estimate the Hurst exponent in data with crossovers and other trend induced effects. Also, this method of detrending the data does not assume a polynomial form for the trend, thus is an improvement over DFA which is prone to crossover effects and other statistical artefacts.

Besides detrending the data, it could also be shown that the Hurst exponents

obtained using FATE can be used to estimate the tail exponent of the distribution of errors. This translates to characterization of the quality of predictions as postulated earlier. We demonstrate this way of error characterization on AL data as the expected tail indices from different models are close to the values estimated by Hill estimator technique.

#### 6.4.2 Microsoft stock price data

The scheme of FATE and the tail exponent estimation from distribution of errors is also applied to data from finance. We obtained the daily opening/closing stock prices of Microsoft (traded as MSFT) from Yahoo finance. This data consists of the stock prices since its first IPO in 1986 till September 2014. The analysis is carried out on the daily closing prices. It is to be noted that we use the actual stock prices for FATE analysis as we want to demonstrate the effect of trends in the actual stock prices as estimation of daily returns is a way of removing the non-stationarities and is undesirable for our analysis.

Results from prediction can be seen in Figure 6.5. Due to no well established attributes that can be used as input for the stock prices other than the usual economic indicators, using which would take the extent of this work beyond our scope, we use only the autonomous model for predictions in this study.

We employed the method of constant updating most recent time steps in formation of trajectory matrix to best utilize information rather than use a constant set of historical values. This method of updating is highly beneficial in systems

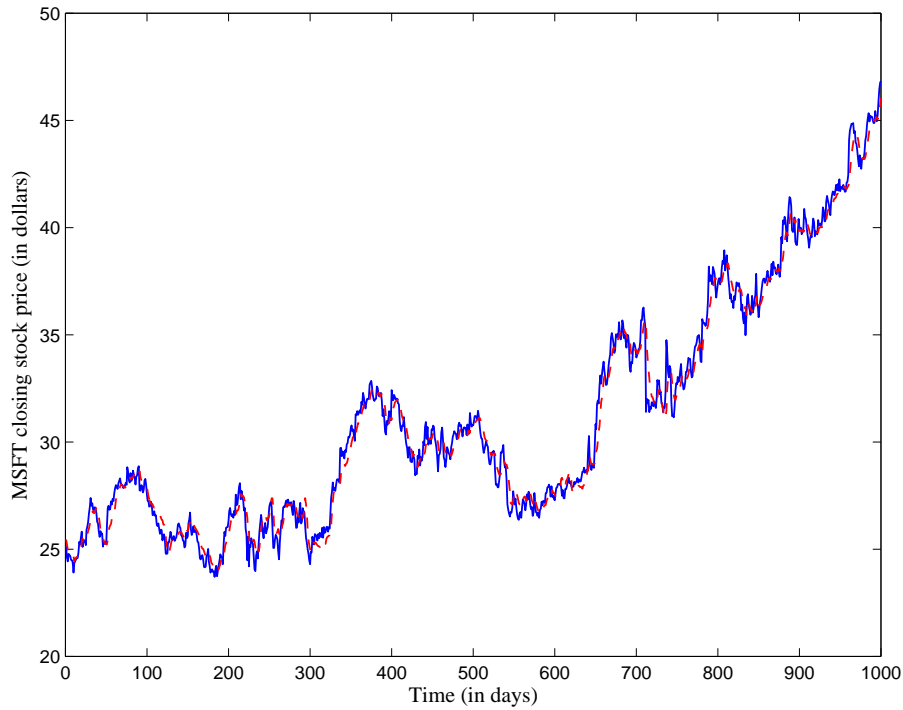


Figure 6.5: MSFT closing stock price (actual in blue) and predicted (in red) for 1000 trading days ending on September 11th, 2014.

that are size limited allowing one to improve the predictions in such cases. In this method, with each update in time series, vectors are updated within the trajectory matrix thus increasing the amount of available phase space information with each update. Single time step predictions were made for 5000 consequent trading days of which predictions from the last 1000 trading days ending on September 11th, 2014 are shown in Figure 6.5.

For the process of phase space reconstruction, we used a time delay of 1 day (shortest time step) and an embedding dimension of 100. The short time delay was due to the short correlation time within the data mainly due to the stochastic nature of the system. With such stochastic systems, increasing embedding dimension is a

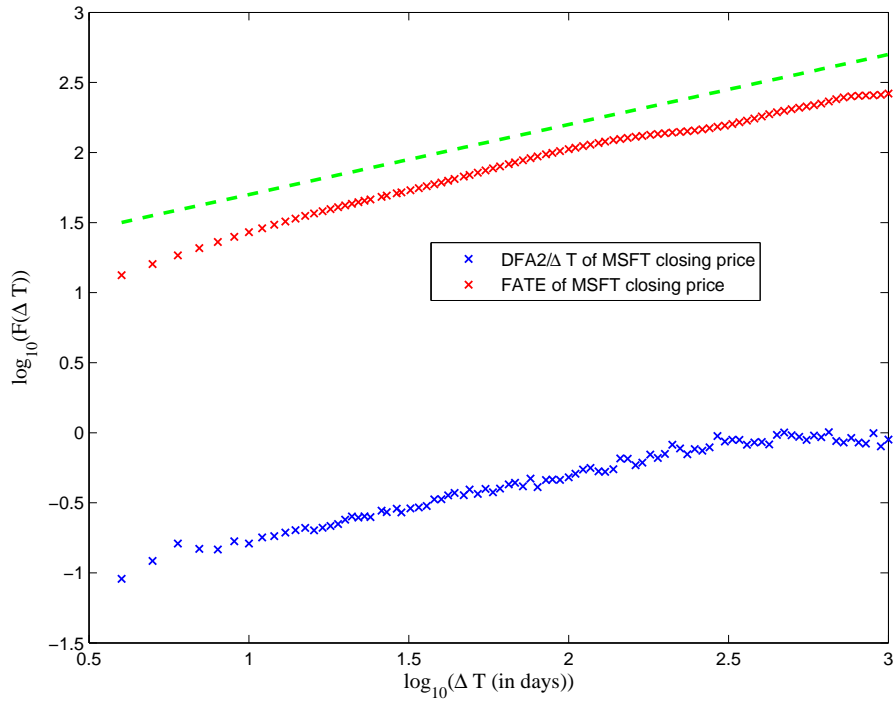


Figure 6.6: FATE applied to MSFT stock price data. The DFA2 of MSFT yields  $\alpha = 1.41$  or  $H = 0.41$ , where as FATE yields  $H = 0.51$ .

necessity as higher complexity implies a higher dimensional system. For our data, we chose this particular dimension taking into account the data size and the necessity of a higher embedding dimension. A weighted mean field filter technique was applied for the prediction of future time steps, with 15 nearest neighbors. Typically increasing the nearest neighbors beyond a certain value would not contribute to accuracy, and in contrary would deteriorate the quality of predictions. The optimization of the simulation parameters such as number nearest neighbors and embedding dimension was performed considering performance of prediction methods over a wide range of choices and the computational efficiency of the algorithm.

Using the predictions, and, the resulting residuals, we perform fluctuation

analysis using the FATE approach. The following results are shown in Figure 6.6. We applied DFA2 technique to estimate the Hurst exponent of actual data due to its non-stationary nature. The Hurst exponent of the daily closing stock price of MSFT is  $0.41 \pm 0.01$  (from the blue plot in Figure 6.6), corresponding to a slight anti-correlated nature, albeit with a non-stationarity of the first order.

FATE analysis utilizing the predictions as trend, yields a Hurst exponent of  $0.51 \pm 0.02$  which shows that the system is near correlation less. Also, if one were to use this value as a measure of the quality of predictions, the expected tail exponent of distribution of errors would be 1.98. The value of the tail exponent using Hill estimator technique is measured to be 2.02. We can note the accuracy of estimation of tail exponent from these two values, one through Hurst exponent and the other through the data.

Results from FATE imply that the predictions utilizing nonlinear techniques are the best possible because the Hurst exponent of errors is  $\approx 0.5$ . This signifies that the errors are near correlationless. Interesting aspect here is that the trends as measured by predictions are separable from the noise or errors. If we compare this to the crossover exhibited in composite indices in Chapter 4, we see that the data from MSFT consists of separable trend and noise similar to NASDAQ, DJIA and N225 indices respectively. Furthermore, the DFA2 of the MSFT index measures the effect of trends within the data as seen by  $\alpha = 1.41$ .

The application of FATE on finance data demonstrates that we have a better technique to detrend data and obtain the Hurst exponent of the noise in data. We also demonstrated the applicability of the Hurst exponent of errors in estimation of

the tail exponent even in presence of the non-stationarities. The Hurst exponent so obtained can be both used as a measure of the level of correlations in the noise within the data and as a measure of predictability of the data utilizing a technique such as nonlinear time series predictions. It is also interesting to note that we could demonstrate through  $H = 0.51$  using FATE that our predictions remove most of the correlations within the system thus yielding noise that is correlationless or void of LRC.

## 6.5 Conclusions

Time series data pertaining to complex systems oftentimes require detrending the data to estimate the Hurst exponent due to non-stationarities from trends within the data. This Hurst exponent aids us in the study of the nature of complexity in such systems as it is measure of LRC in noise within such systems. However, due to existence of atypical trends, process of detrending is complicated and riddled with biases. Our proposed technique detrends such systems *a priori* and due to lack of assumptions of the nature of trends, is void of any statistical artefacts. This method has been demonstrated on a theoretical model with additive trends of periodic nature such as correlated gaussian noise added to a sinusoidal trend. DFA is unable to remove the effect of the trend in such cases and often times, the range of estimation of the Hurst exponent which we refer to as the operating range is quite limited as shown in Figure 6.1. FATE provides a way to address such trends *a priori* with no assumptions.

Also, prediction of complex systems is complicated due to noise or high frequency aspect within the power spectrum of their data. Understanding their predictability thus is essential to improve the technique of prediction or model them better. Characterization of predictions using a mean square estimate of the residuals (or errors) is simplistic and ignores the existence of heavy tails in their distribution. Also, such a method of characterization may not yield information with respect to LRC nature of the noise due to which heavy tails occur. Utilizing the existence of a well known Taqqu's theorem that relates the Hurst exponent to the tail index of heavy tailed distributions, we propose to use the Hurst exponents estimated after applying FATE on data to characterize the error. This method of characterizing error, not only yields more information about the quality of predictions, but also helps estimate a tail index in cases where such estimation is complicated due to many reasons. These ideas are applied to data from space weather i.e., AL index and finance i.e., daily stock closing prices for Microsoft.

Our analysis yields that we can successfully detrend the data and obtain Hurst exponents in cases with non-stationarities. We also demonstrate how these Hurst exponents are used to estimate the tail exponent. The estimated tail exponents are seen to be very close to the ones estimated from a well know Hill estimator technique. Further work is necessary to demonstrate how this scheme would improve tail exponent estimation in multi-scale data.



## Chapter 7: Summary

Fluctuation analysis is being widely applied to uncover long range correlation structure within various time series data. Complex systems often demonstrate multi-scale nature, particularly in time, due to interactions between distributed systems at different scales. Identification of crossover in fluctuation functions has been discovered previously in heart beat variability data, more recently in volatility data from finance and most recently in space weather index data due to our work.

Existence of such scaling crossover aids in characterizing the multi-scale nature of data. Taking advantage of this, we propose a regression scheme to model two asymptotic scales with a scale mixed regime. Application of this scheme on aforementioned data yielded valuable new results for each case. This is one of the two applications of fluctuation analysis shown in this dissertation.

Predictability of data is effected by heavy tailed distribution of noise in it. Self similarity has a propensity to cause extreme events due to build up by correlations. An improved scheme to estimate the Hurst exponent, a measure of self similarity or long range correlations, is proposed and applied on our data. Hurst exponent obtained from this scheme can be related to the tail exponent-a measure of fat tailedness in probability distribution. Hence, the second application of fluctuation analysis

includes characterizing predictability of complex systems using Hurst exponent of noise in such systems.

We explain in greater detail the importance of work, results presented in this dissertation, and, propose a possible course for future work in each case.

### **1. Characterizing DFA of mBm data**

Detrended Fluctuation Analysis (DFA) is a valuable and widely used tool for fluctuation analysis in non-stationary data. More practical systems have a time varying Hurst exponent structure as the concept of self similarity does not assure homogeneity in scales at all times. Multifractional Brownian motion (mBm) is a theoretical model that incorporates such time varying Hurst exponents. Monte Carlo methods simulating ensembles of mBm with DFA applied to them helps us understand its performance as a technique to estimate exponents in a sliding window fashion, as it is most commonly applied. This serves as an introduction to DFA and sets benchmarks on its robustness [Setty and Sharma 2015].

We also present our motivation for choosing mBm model due to the observed time varying Hurst exponents in AL index data from space weather. While we do our best to present a case for DFA as a sliding window estimator, there were some unanswered questions with regards to results seen in this work. Particularly, improved performance of DFA in cases where the data is predominantly self similar i.e., the Hurst exponent is greater than 0.7 for most of the time is unexplained. Future work would require explanation for this phenomenon. Also, one needs to correlate the time averaged Hurst exponent within a window with other factors from data such as its seasonality or other relevant factors. This would be a valuable

step in explaining time varying nature of the exponents in mBm like data.

## **2. Scaling crossover analysis in space weather, physiology and finance**

Identification and measurement of the crossover phenomenon is critical to understand multi-scaled nature in data from complex systems. We selected data from varied physical systems that all demonstrate two scales asymptotically. A hyperbola has two asymptotes and an intermediate curved region. Regressing fluctuation functions with a single crossover onto a hyperbola, hence, is the logical method for scale exponent estimation in such systems.

Space weather data is known to be multi-scaled from previous work. Taking cue from this, we applied fluctuation analysis and proved the existence of crossover in data for the first time. Furthermore, using hyperbolic regression, we could measure the asymptotic scaling exponents before and after the crossover time. However, variance is offset when DFA is applied to measure fluctuation functions due to the detrending operation. Hence, a better method is required to estimate crossover time scale in our data. This necessitated proposal of a model for crossover phenomenon i.e., stochastic OU-Langevin and using theoretical variance from this model to estimate the crossover time. In this work, we demonstrate how crossover can be characterized, and, how modeling the crossover can help us measure important features pertaining to its multi-scaled nature. Further work is required in the direction of stochastic modeling of AL data, particularly, attention is required to better understand the nature of stochasticity in AL data. Fractional Gaussian noise with similar Hurst exponent as that of our AL data fails to explain the sudden bursts or

intermittency seen in the data. We need to find an alternative model that captures scaling crossover feature whilst explaining the bursting behavior.

Generalized volatility data from composite stock indices are known to exhibit crossover(s) in fluctuation analysis. The extent of crossover can be controlled by a weight factor from the definition of generalized volatility. However, generalized volatility of indices with same weight factors have been found to show opposing features with respect to concavity and convexity in the crossover. It is interesting to note that this property has a globally varying aspect owing to factors that impact the index at a particular geography. This is one of the first such analysis performed with respect to varying nature of markets. Due to interesting aspects discovered from our analysis, this study should be expanded to include more global indices besides the five indices used in our work. Characterization of financial markets using crossover in fluctuation analysis is a novel idea and further work is required in this direction.

Heart rate variability data exhibits multiple time scales due to competition between sympathetic and parasympathetic nervous systems to control cardiac pacemaker cells. We applied hyperbolic regression to such scaling crossover and studied the results closely. Very interesting ideas emerged due to this study. A new parameter that measures the extent of crossover (EOC) is proposed. What is striking about this parameter is that there is a wide separation in EOC values between NSR and CHF cases. Further work is required in this direction to better characterize the EOC parameter in many more patients. It would involve a blind study to test if EOC is effective in differentiating CHF from NSR cases. Furthermore, being even

more ambitious, we feel that there is a correlation between magnitude of EOC and severity of CHF condition in the patient which needs to be better understood. We feel that increased EOC may imply higher intensity of the ailment adding further value to the utility of EOC parameter.

### **3. Characterizing predictability using Hurst exponents**

Trends and seasonality seen in empirical data make estimation of Hurst exponent difficult. Although DFA estimates Hurst exponents from non-stationary data, it does not remove the trend as claimed. The resulting fluctuation functions are affected, particularly when the trends are not polynomial addressable. As a solution to this problem, we propose removing trends beforehand using nonlinear time series predictions and label this method as Fluctuation Analysis after Trend Elimination (FATE). This method is applied on a test case and its success is quite evident.

We refer to that what cannot be predicted or the residual after removing what can be predicted as noise in systems. Certain types of noise with heavy-tail distributions have higher probability of extreme event occurrence in them and hence in systems containing them. Hence, the predictability of data that contain these heavy-tailed noise as variability suffers. Long range correlations (LRC) in noise can cause extreme events because of buildups by correlations. Hurst exponent is a measure of LRC. Hence, it is logical to use Hurst exponent of fluctuations in systems to characterize their predictability and forms the basis for our work. We estimate Hurst exponents using FATE on data from space weather and finance. Relationship between Hurst exponent and tail exponent from probability distribution is well established. Using this relationship, we can relate the tail exponent measured by

a well known technique called Hill estimator to the Hurst exponents obtained by procedure mentioned earlier proving the validity of our idea.

Hill estimator, although to a large extent is successful in measuring tail exponents from empirical data is prone to many issues. One of them is non-convergence in the series leading to the tail exponent. This and other such aberrations are comically referred to as Hill horror plots. “If it did not happen yet, it cannot be measured” can be used to describe the problem, particularly for extreme events. LRC and evidence of LRC is present in the system and can be seen even without the existence of extreme events in the data. Further work is necessary to prove the ability of obtaining tail exponent from Hurst exponent of finite small data sets. This is an essential step to solve problems in tail exponent estimation particularly from techniques such as the Hill estimator.

## Bibliography

- H. D. Abarbanel, R. Brown, J. J. Sidorovich, and T. S. Tsimring. The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.*, 65:1331, 1993.
- P. Abry and F. Sellan. The Wavelet-Based Synthesis for Fractional Brownian Motion: Remarks and Fast Implementation. *Applied and Computational Harmonic Analysis*, 3(4):377–383, 1996.
- V. V. Anh, J. M. Yong, and Z. G. Yu. Stochastic modeling of the auroral electrojet index. *Journal of Geophysical Research*, 113:A10215, 2008.
- D. N. Baker, A. J. Klimas, R. L. McPherron, and J. Buchner. The evolution from weak to strong geomagnetic activity: An interpretation in terms of deterministic chaos. *Geophys. Res. Lett.*, 17:41, 1990.
- D. N. Baker, A. J. Klimas, D. Vassiliadis, and T. I. Pulkkinen. Reexamination of driven and unloading aspects of magnetospheric substorms. *Journal of Geophysical Research*, 102(A4):7169–7177, April 1997.
- D. N. Baker, J. H. Allen, S. G. Kanekal, and G. D. Reeves. Complex Systems Modeling: Using Metaphors From Nature in Simulation and Scientific Models. *Eos, Transactions American Geophysical Union*, 79(40):477–483, 1998.
- Y. Bar-Yam. General Features of Complex Systems. *Encyclopedia of Life Support Systems*, 2002.
- J. Barunik, T. Aste, T. Di. Matteo, and R. Liu. Understanding the source of multifractality in financial markets. *Physica A*, 391(17):4234–4251, 2012.
- J. Beran. *Statistics for Long-Memory Processes*. CRC Press, 1994.
- P. R. Bertrand, H. Abdelkader, and K. Samia. Modelling NASDAQ series by sparse Multifractional Brownian Motion. *Methodol Comput Appl Probab*, 14:107–124, 2012.
- M. Brennan, M. Palaniswami, and P. Kamen. Do existing measures of poincaré plot geometry reflect non-linear features of heart rate variability? *IEEE Transactions on Biomedical Engineering*, 48:1342–1347, 2001.

- R. M. Bryce and K. B. Sprague. Revisiting detrended fluctuation analysis. *Scientific Reports*, 2:315, 2012.
- S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, M. E. Matsu, C.-K. Peng, M. Simons, and H. E. Stanley. Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. *Phys. Rev. E*, 51:5084–5091, May 1995.
- A. Carbone, Castelli G., and Stanley H. E. Time-dependent Hurst exponent in financial time series. *Physica A*, 334(1):267–271, 2004.
- P. G. Carles. Empirical evidence of long-range correlations in stock returns. *Physica A*, 287:396, 2000.
- G. Chan and A. T. A. Wood. Simulation of Multifractional Brownian Motion. *Proc. COMPSTAT'98*, pages 233–238, 1998.
- S. Chandrasekhar. Stochastic Problems in Physics and Astronomy. *Rev. Mod. Phys.*, 15:1–89, 1943.
- J. Chen and A. S. Sharma. Modeling and prediction of the magnetospheric dynamics during intense geospace storms. *J. Geophys. Res.*, 111, 2006.
- A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. *SIAM Rev.*, 51:661–703, 2009.
- J.-F. Coeurjolly. *Statistical inference for fractional and multifractional Brownian motions*. PhD thesis, University Joseph Fourier, Grenoble, France., 2000.
- T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., New York, 1991.
- M. Cristelli. *Complexity in Financial Markets*. Springer Theses, 2014.
- P. J. Davis. *Circulant Matrices*. Wiley, New York, 1970.
- T. N. Davis and M. Sugiura. Auroral electrojet activity index AE and its universal time variations. *Journal of Geophysical Research*, 71(3):785–801, 1966.
- A. Echelard, O. Barriere, and J. Lévy-Véhel. Terrain modeling with multifractional Brownian motion and self-regulating processes. *Computer Vision and Graphics*, 6374:342–351, 2010.
- C. Eom, S. Choi, G. Oh, and W.-S. Jung. Hurst exponent and prediction based on weak-form efficient market hypothesis of stock markets. *Physica A*, 387:4630–4636, 2008.
- A. M. Fraser and H. L. Swinney. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A.*, 33:1134–1140, 1986.



- S. Gaci and N. Zaourar. Heterogeneities characterization from velocity logs using multifractional Brownian motion. *Arab. J. Geosci.*, 4(3-4):535–541, 2011.
- R.G. Gallager. *Information Theory and Reliable Communication*. John Wiley and Sons, Inc., New York, 1968.
- A. L. Goldberger. Is the normal heartbeat chaotic or homeostatic? *News Physiol Sci.*, pages 87–91, 1991.
- A. L. Goldberger. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. *The Lancet*, 347:1312–1314, 1996.
- A. L. Goldberger, D. R. Rigney, and B. J. West. Chaos and fractals in human physiology. *Sci. Am.*, 262:42–49, 1990.
- D. G. Goodin, P. A. Fay, and M. J. McHugh. Climate variability at multiple time scales: implications for productivity in tallgrass prairie. 2002.
- E. A. Guggenheim. The Principle of Corresponding States. *J. Chem. Phys.*, 13:253, 1945.
- S. Havlin, L. A. Amaral, Y. Ashkenazy, A. L. Goldberger, I. PCh, C. K. Peng, and H. E. Stanley. Application of statistical physics to heartbeat diagnosis. *Physica A*, 274:99–110, 1999.
- B. M. Hill. A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3:1163–1174, 1975.
- L. Hongre, P. SAILHAC, M. Alexandrescu, and J. Dubois. Nonlinear and multifractal approaches of the geomagnetic field. *Physics of the Earth and Planetary Interiors*, 110(3–4):157 – 190, 1999.
- K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley. Effect of Trends on Detrended Fluctuation Analysis. *Phys. Rev. E.*, 64:011114, 2001.
- H. E. Hurst. Long-term storage capacity of reservoirs (with discussion). *Tran. Am. Soc. Civ. Eng.*, 116:770–808, 1951.
- H. E. Hurst, R. P. Black, and Y. M. Simaikhah. *Long-term storage, an experimental study*. Constable, London, UK, 1965.
- M. Ignaccolo, M. Latka, and B. J. West. Detrended fluctuation analysis of scaling crossover effects. *EPL (Europhysics Letters)*, 90(1):10009, 2010.
- A. Kamont. On the fractional anisotropic Wiener field. *Probab. Math. Statist.*, 16(1):85–98, 1996.
- J.W. Kantelhardt, E. Kpścielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde. Detecting Long-range Correlations with Detrended Fluctuation Analysis. *Physica A*, 295:441–454, 2001.

- J.W. Kantelhardt, S. A. Zschiegner, E. Kpścielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley. Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series. *Physica A*, 316:87–114, 2002.
- J.K. Kanters, N. H. Holstein-Rathlou, and E. Agner. Lack of evidence for low-dimensional chaos in heart rate variability. *Journal of Cardiovascular Electrophysiology*, 5:591–601, 1994.
- S. Karlin and V. Brendel. Patchiness and correlations in DNA sequences. *Science*, 259:677–680, 1993.
- M. G. Kivelson and C. T. Russell. *Introduction to Space Physics*. Cambridge University Press, 1995.
- R. E. Kleiger, J. P. Miller, J. T. Bigger, and A. J. Moss. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am J Cardiol*, 59:256–262, 1987.
- E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H. Schellnhuber. Indication of a Universal Persistence Law Governing Atmospheric Variability. *Phys. Rev. Lett.*, 81:729–732, 1998.
- P. Langevin. On the Theory of Brownian Motion. *C. R. Acad. Sci. (Paris)*, 146:530–533, 1908.
- B. LeBaron. Time scales, Agents, and Empirical Finance. *Medium Econometricshce*, 14(3):20–25, 2006.
- W. E. Leland, M. S. Taqqu, and W. Willinger. On the self-similar nature of ethernet traffic. *ACM/IEEE Transactions networking*, 2:1–15, 1994.
- S. Lennartz and A. Bunde. Distribution of natural trends in long-term correlated records: A scaling approach. *Phys. Rev. E*, 84:021129, 2011.
- M. N. Levy. Sympathetic-parasympathetic interactions in the heart. *Circ. Res.*, 29:437–445, 1971.
- M. Li and W. Zhao. Quantitatively investigating locally weak stationarity of modified multifractional Gaussian noise. *Physica A*, 391(24):6268–6278, 2012.
- F. Lillo and J. D. Farmer. The Long Memory of the Efficient Market. *Studies in Nonlinear Dynamics & Econometrics*, 8, 2004.
- Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley. The Statistical Properties of the Volatility of Price Fluctuations. *Phys. Rev. E*, 60:1390–1400, 1999.
- E.N. Lorenz. Determining nonperiodic flow. *J. Atmos. Sci.*, 20:130–141, 1963.

- B. Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36:394–419, 1963.
- B. Mandelbrot. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer, 1997.
- B. B. Mandelbrot and J. R. Wallis. Some long-run properties of geophysical records. *Water Resour. Res.*, 5:321–340, 1969.
- B.B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.
- R.H. McCuen. *Statistical Methods for Engineers*. Prentice Hall, Englewood Cliffs, NJ, 1985.
- T. A. McDonagh, R. S. Gardner, A. L. Clark, and H. Dargie. *Oxford Textbook of Heart Failure*. Oxford University Press, 2011.
- M. Mitzenmacher. A Brief History of Generative Models for Power Law and Log-normal Distributions. *Internet Mathematics*, 1:226–251, 2004.
- M. S. Movahed, G. R. Jafari, F. Ghasemi, S. Rahvar, and M. R. Tabar. Multifractal detrended fluctuation analysis of sunspot time series. *Journal of Statistical Mechanics: Theory and Experiment*, 0602:003, 2006.
- M. E. J. Newman. Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics*, 46:323–351, 2005.
- P. O’Leary and P. Zsombor-Murray. Direct and specific least-square fitting of hyperbolæ and ellipses. *Journal of Electronic Imaging*, 13:492, 2004.
- N. H. Packard, J. P. Crutchfeld, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Phys. Rev. Lett.*, 45:712–716, 1980.
- R.F. Peltier and Lévy-Véhel. Multifractional Brownian motion : definition and preliminary results. *Rapport de recherche de l’INRIA*, 2645, 1995.
- C.-K. Peng, S. V. Buldyrev, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley. Long-range correlations in nucleotide sequences. *Nature*, 356:168–170, 1992.
- C. K. Peng, J. E. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, and A. L. Goldberger. Long-Range Anticorrelations and Non-Gaussian Behavior of the Heart-beat. *Phys Rev Lett.*, 70:1343–1346, 1993.
- C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic Organization of DNA Nucleotides. *Phys. Rev. E*, 49:1685–1689, 1994.

- C. K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger. Quantification of Scaling Exponents and Crossover Phenomena in Nonstationary Heartbeat Time Series. *Chaos*, 5:82–87, 1995.
- C. K. Peng, J. M. Hausdorff, S. Havlin, J. E. Mietus, H. E. Stanley, and A. L. Goldberger. Multiple-time scales analysis of physiological time series under neural control. *Physica A*, 249:491–500, 1998.
- J. A. Ratcliffe. *An Introduction to the Ionosphere and Magnetosphere*. CUP Archive, 1972.
- S. I. Resnick. Heavy tail modeling and teletraffic data. *The Annals of Statistics*, 25: 1805–1869, 1997.
- S. I. Resnick and C. Starica. Smoothing the Hill estimator. *Adv. Appl. Prob.*, 29: 271–293, 1997.
- L. C. Rogerio and G.L. Vasconcelos. Long-range correlations and nonstationarity in the Brazilian stock market. *Physica A*, 329(1-2):231–248, 2003.
- D. Ruelle. *Chaotic Evolution and Strange Attractors*. Cambridge Press, Cambridge, UK, 1989.
- D. Rybski and A. Bunde. On the detection of trends in long-term correlated records. *Physica A*, 388:1687–1695, 2009.
- M. Rypdal and K. Rypdal. Stochastic modeling of the AE index and its relation to fluctuation in  $B_z$  of the IMF on time scales shorter than substorm duration. *Journal of Geophysical Research*, 115:A11216, 2010.
- G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, 1994.
- C. Schäfer, M. G. Rosenblum, J. Kurths, and H. H. Abel. Heartbeat synchronized with ventilation Heartbeat synchronized with ventilation. *Nature*, 392:239–240, 1998.
- F. Schmitt, D. Schertzer, and S. Lovejoy. Multifractal analysis of foreign exchange data. *Applied Stochastic Models and Data Analysis*, 15:29–53, 1999.
- V. A. Setty and A. S. Sharma. Characterizing Detrended Fluctuation Analysis of Multifractional Brownian Motion. *Physica A*, 419:698–706, 2015.
- C. E. Shannon. The mathematical theory of communication. *Bell Syst. Techn. Journal*, 27:379–423, 1948.
- A. S. Sharma. Assessing the Magnetosphere’s Nonlinear Behavior: Its Dimension is Low, its Predictability High. *Rev. Geophys.*, 33:645–650, 1995.

- A.S. Sharma and T. Veeramani. Extreme events and long-range correlations in space weather. *Nonlin. Processes Geophys.*, 18:719–725, 2011.
- M. Shelhamer. *Nonlinear Dynamics in Physiology: A State-space Approach*. World Scientific, Singapore, 2006.
- H. Sheng, Y. Chen, and T. Qiu. Tracking performance of Hurst Estimators for multifractional Gaussian processes. In *Proceedings of FDA'10. The 4th IFAC Workshop Fractional Differentiation and its Applications*, 2010.
- M. Shi-Hao. Crossover Phenomena in Detrended Fluctuation Analysis Used in Financial Markets. *Commun. Theor. Phys.*, 51:358–362, 2009.
- D. Sornette. *Critical Phenomena in Natural Sciences*. Springer, Berlin, 2006.
- H. E. Stanley, S. V. Buldyrev, A. L. Goldberger, Z. D. Goldberger, S. Havlin, R. N. Mantegna, S. M. Ossadnik, C.-K. Peng, and M. Simons. Statistical mechanics in biology: how ubiquitous are long-range correlations? *Physica A*, 205:214–53, 1994.
- S. A. Stoev, G. Michailidis, and M. S. Taqqu. Estimating Heavy-Tail Exponents Through Max Self-Similarity. *IEEE TRANSACTIONS ON INFORMATION THEORY*, 57:1615–1636, 2011.
- R. J. Storella, H. W. Wood, and K. M. Mills. Approximate entropy and point correlation dimension of heart rate variability in healthy subjects. *Integrative Physiological Behavioral Science*, 33:315–320, 1994.
- J. V. Sutcliffe. Obituary: Harold Edwin Hurst. *Hydrolog Sci J*, 24:539–541, 1979.
- J. Takalo and J. Timonen. Characteristic time scale of auroral electrojet data. *Geophysical Research Letters*, 21:617–620, 1994.
- F. Takens. *Detecting strange attractors in turbulence, in Dynamical Systems and Turbulence*. Springer, Berlin, Germany, 1981.
- N. N. Taleb. *The Black Swan*. Random House and Penguin, 2007.
- M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: An empirical study. *Fractals*, 3:785–798, 1995.
- M. S. Taqqu, W. Willinger, and R. Sherman. Proof of a fundamental result in self-similar traffic modeling. *ACM SIGCOMM Computer Communication Review*, 27: 5–23, 1997.
- B. T. Tsurutani, M. Sugiura, T. Iyemori, B. E. Goldstein, W. D. Gonzalez, S. I. Akasofu, and E. J. Smith. The nonlinear response of AE to the imf  $B_s$  driver: A spectral break at 5 hours. *Geophysical Research Letters*, 17(3):279–282, March 1990.

- G. E. Uhlenbeck and L. S. Ornstein. On the Theory of the Brownian Motion. 36: 823–841, 1930.
- A. Y. Ukhorskiy, M. I. Sitnov, A. S. Sharma, and K. Papadopoulos. Global and multiscale aspects of magnetospheric dynamics in local-linear filters. *J. Geophys. Res.*, 107:2369, 2002.
- J. A. Valdivia, A. S. Sharma, and K. Papadopoulos. Prediction of magnetic storms by nonlinear models. *Geophys. Res. Lett.*, 23:2899, 1996.
- D. V. Vassiliadis, A. S. Sharma, T. E. Eastman, and K. Papadopoulos. Low Dimensional chaos in magnetospheric activity from AE time series. *Geophys. Res. Lett.*, 17:1841, 1990.
- M. C. Wang and G. E. Uhlenbeck. On the Theory of the Brownian Motion II. *Rev. Mod. Phys.*, 17:323–342, 1945.
- J.A. Wanliss and P. Dobias. Space Storm as a Phase Transition. *J. Atmos. Sol. Terr. Phys.*, 69:675–684, 2007.
- J. A. Wanliss and D. O. Cersosimo. Scaling properties of high latitude magnetic field data during different magnetospheric conditions. In *International Conference Substorms*, pages 325–329, 2006.
- H. L. Wei, S. A. Billings, and M. Balikhin. Analysis of the geomagnetic activity of the Dst index and self-affine fractals using wavelet transforms. *Nonlinear Processes in Geophysics*, 11(3):303–312, 2004.
- K. G. Wilson. Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture. *Phys. Rev. B*, 4:3174, 1971.
- A. T. A. Wood and G. Chan. Simulation of stationary Gaussian processes in  $[0, 1]^d$ . *J. Comp. Graph. Statist.*, 3(4):409–432, 1994.